

**PRIORITY
DOCUMENT**
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)



REC'D 03 OCT 2000
WIPO PCT

09/830902

BREVET D'INVENTION

FR 00/02432

CERTIFICAT D'UTILITÉ - CERTIFICAT D'ADDITION

E JV

COPIE OFFICIELLE

Le Directeur général de l'Institut national de la propriété industrielle certifie que le document ci-annexé est la copie certifiée conforme d'une demande de titre de propriété industrielle déposée à l'Institut.

Fait à Paris, le **29 AOUT 2000**

Pour le Directeur général de l'Institut
national de la propriété industrielle
Le Chef du Département des brevets

Martine PLANCHE

INSTITUT
NATIONAL DE
LA PROPRIÉTÉ
INDUSTRIELLE

SIEGE
26 bis. rue de Saint Petersburg
75800 PARIS Cédex 08
Téléphone : 01 53 04 53 04
Télécopie : 01 42 93 59 30

REQUÊTE EN DÉLIVRANCE

26 bis, rue de Saint Pétersbourg
75800 Paris Cedex 08

Téléphone : 01 53 04 53 04 Télécopie : 01 42 93 59 30

Confirmation d'un dépôt par télécopie ☐

Cet imprimé est à remplir à l'encre noire en lettres capitales

<p>DATE DE REMISE DES PIÈCES 03.09.99</p> <p>N° D'ENREGISTREMENT NATIONAL 9911097</p> <p>DÉPARTEMENT DE DÉPÔT 75</p> <p>DATE DE DÉPÔT 03/09/99</p>		<p>1 NOM ET ADRESSE DU DEMANDEUR OU DU MANDATAIRE À QUI LA CORRESPONDANCE DOIT ÊTRE ADRESSÉE</p> <p>CABINET REGIMBEAU 26, Avenue Kléber 75116 PARIS</p>									
<p>2 DEMANDE Nature du titre de propriété industrielle</p> <p><input checked="" type="checkbox"/> brevet d'invention <input type="checkbox"/> demande divisionnaire</p> <p><input type="checkbox"/> certificat d'utilité <input type="checkbox"/> transformation d'une demande de brevet européen</p> <p><input type="checkbox"/> brevet d'invention <input type="checkbox"/> certificat d'utilité n°</p> <p>Établissement du rapport de recherche <input type="checkbox"/> différé <input checked="" type="checkbox"/> immédiat</p> <p>Le demandeur, personne physique, requiert le paiement échelonné de la redevance <input type="checkbox"/> oui <input type="checkbox"/> non</p> <p>Titre de l'invention (200 caractères maximum)</p>		<p>n° du pouvoir permanent 238012 D18374 NIP références du correspondant 01 45 00 92 02 téléphone</p> <p>date</p>									
<p>CLONAGE, EXPRESSION ET CARACTERISATION DU GENE SPG4 RESPONSABLE DE LA FORME LA PLUS FREQUENTE DE PARAPLEGIE SPASTIQUE AUTOSOMIQUE DOMINANTE.</p>											
<p>3 DEMANDEUR (S) n° SIREN</p> <p>Norm et prénoms (souligner le nom patronymique) ou dénomination</p> <p>CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS)</p>		<p>code APE-NAF</p> <p>Forme juridique</p> <p>ETABLISSEMENT PUBLIC A CARACTERE SCIENTIFIQUE ET TECHNO...</p>									
<p>Nationalité (s) Française</p> <p>Adresse (s) complète (s)</p> <p>3, rue Michel Ange, 75016 PARIS</p>		<p>Pays FR</p>									
<p>4 INVENTEUR (S) Les inventeurs sont les demandeurs <input type="checkbox"/> oui <input checked="" type="checkbox"/> non Si la réponse est non, fournir une désignation séparée</p>											
<p>5 RÉDUCTION DU TAUX DES REDEVANCES <input type="checkbox"/> requise pour la 1ère fois <input type="checkbox"/> requise antérieurement au dépôt ; joindre copie de la décision d'admission</p>											
<p>6 DÉCLARATION DE PRIORITÉ OU REQUÊTE DU BÉNÉFICE DE LA DATE DE DÉPÔT D'UNE DEMANDE ANTÉRIEURE</p> <table border="1"> <thead> <tr> <th>pays d'origine</th> <th>numéro</th> <th>date de dépôt</th> <th>nature de la demande</th> </tr> </thead> <tbody> <tr> <td> </td> <td> </td> <td> </td> <td> </td> </tr> </tbody> </table>				pays d'origine	numéro	date de dépôt	nature de la demande				
pays d'origine	numéro	date de dépôt	nature de la demande								
<p>7 DIVISIONS antérieures à la présente demande n° date n° date</p>											
<p>8 SIGNATURE DU DEMANDEUR OU DU MANDATAIRE (nom et qualité du signataire)</p> <p>321169 <i>[Signature]</i></p>		<p>SIGNATURE DU PRÉPOSÉ À LA RÉCEPTION</p> <p>SIGNATURE APRÈS ENREGISTREMENT DE LA DEMANDE À L'INPI</p> <p><i>[Signature]</i></p>									

DÉSIGNATION DE L'INVENTEUR

(si le demandeur n'est pas l'inventeur ou l'unique inventeur)

DEPARTEMENT DES BREVETS

26bis, rue de Saint-Petersbourg

75800 Paris Cédex 08

Tél. : 01 53 04 53 04 - Télécopie : 01 42 93 59 30

N° D'ENREGISTREMENT NATIONAL

99 11097

**TITRE DE L'INVENTION : CLONAGE, EXPRESSION ET CARACTERISATION DU GENE SPG4
RESPONSABLE DE LA FORME LA PLUS FREQUENTE DE PARAPLEGIE SPASTIQUE
AUTOSOMIQUE DOMINANTE.**

LE(S) SOUSSIGNÉ(S)

**CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS)
3, rue Michel Ange, 75016 PARIS**

DÉSIGNE(NT) EN TANT QU'INVENTEUR(S) (indiquer nom, prénoms, adresse et souligner le nom patronymique) :

**WEISSENBACH Jean
163, rue de Vaugirard
75015 Paris, FR**

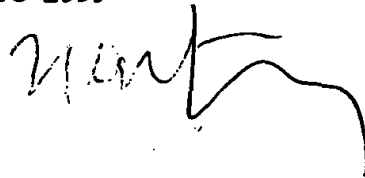
**HAZAN Jamilé
52, avenue René Coty
75014 PARIS, FR**

NOTA : A titre exceptionnel, le nom de l'inventeur peut être suivi de celui de la société à laquelle il appartient (société d'appartenance) lorsque celle-ci est différente de la société déposante ou titulaire.

Date et signature (s) du (des) demandeur (s) ou du mandataire

3 septembre 1999

CABINET REGIMBEAU



921169

DOCUMENT COMPORTANT DES MODIFICATIONS

PAGE(S) DE LA DESCRIPTION OU DES REVENDICATIONS OU PLANCHE(S) DE DESSIN			R.M.*	DATE DE LA CORRESPONDANCE	TAMPON DATEUR DU CORRECTEUR
Modifiée(s)	Supprimée(s)	Ajoutée(s)			
p. 43				31.03.00	04 AVR. 2000 - VD
p. 124-126			X	31.03.00	04 AVR. 2000 - VD
p. 123-127		p. 128-135	X	31.03.00	04 AVR. 2000 - VD
p. 132			X	10.04.00	13 AVR. 2000 - VD

Un changement apporté à la rédaction des revendications d'origine, sauf si celui-ci découle des dispositions de l'article R.612-36 du code de la Propriété Intellectuelle, est signalé par la mention «R.M.» (revendications modifiées).

ORIGINAL

CLONAGE, EXPRESSION ET CARACTERISATION DU GENE SPG4 RESPONSABLE DE LA FORME LA PLUS FREQUENTE DE PARAPLEGIE SPASTIQUE AUTOSOMIQUE DOMINANTE.

5 L'invention concerne l'identification et la caractérisation du gène SPG4 codant pour la spastin, responsable de la forme la plus fréquente de paraplégie spastique familiale (PSF) autosomique dominante, le clonage et la caractérisation de son ADNc ainsi que les polypeptides correspondants. L'invention concerne également des vecteurs, des cellules transformées et des animaux transgéniques ainsi que des
10 méthodes de diagnostic et des méthodes de sélection d'un composé chimique ou biochimique capable d'interagir directement ou indirectement avec un polypeptide selon l'invention.

Les paraplégies spastiques familiales (PSF) sont des affections dégénératives du système nerveux central caractérisées par une spasticité bilatérale et progressive
15 des membres inférieurs. Elles se manifestent cliniquement par des difficultés de la marche pouvant évoluer en une paralysie totale des deux jambes. La physiopathologie de cet ensemble de maladies est à ce jour peu documentée ; toutefois, les données anatomopathologiques permettent de conclure que l'atteinte est limitée aux faisceaux pyramidaux responsables de la motricité volontaire dans la moelle épinière (1). Il existe
20 différentes formes cliniques et génétiques de PSF. On distingue cliniquement les PSF dites « pures » correspondant à une spasticité isolée des membres inférieurs, d s PSF « complexes » pour lesquelles la spasticité des jambes est associée à d'autres manifestations cliniques de type neurologique ou non (2). D'un point de vue génétique, les PSF peuvent être transmises selon le mode autosomique dominant (PSF-AD),
25 autosomique récessif (PSF-AR) ou lié au chromosome X (PSF-X). La forme « pure » de PSF qui est le plus souvent transmise selon le mode autosomique dominant demeure la plus fréquente (environ 80 % des PSF) (1). L'incidence des PSF qui reste difficile à estimer en raison des rares études épidémiologiques et de l'importante variabilité clinique varie de 0,9 : 100 000 au Danemark, 3 à 9,6 : 100 000 dans
30 certaines régions d'Espagne (4) ou 14 : 100 000 en Norvège (5) (environ 3 : 100 000 en France).

En plus de cette grande variabilité clinique qui est observée non seulement entre les différentes familles mais aussi entre différents membres atteints d'une même famille, les PSF se caractérisent également par une importante hétérogénéité
35 génétique. Dans le cas des PSF-AD, quatre loci ont été identifiés à ce jour sur les

chromosomes 14 (locus SPG3) (6), 2 (locus SPG4) (7, 8), 15 (locus SPG6) (9) et 8 (locus SPG8) (10). L'étude d'un grand nombre de familles présentant une PSF-AD a montré que le gène porté par le chromosome 2 est un locus majoritaire de cette forme de la maladie, retrouvé dans 40 à 50 % des familles analysées (11, 12). Un phénomène d'anticipation a été observé dans certaines familles de PSF liée au locus SPG4 ; ce phénomène a par la suite été associé à l'expansion d'une répétition (CAG)_n mise en évidence dans 6 familles danoises (13) par la technique de RED (pour Rapid Expansion Detection). Cette expansion n'a cependant jamais pu être confirmée dans aucune des familles testées par cette même méthode ni par la recherche systématique de séquences de type (CAG)_n dans les cartes physiques composées de clones YAC (pour Yeast Artificial Chromosome) ou BAC (pour Bacterial Artificial Chromosome) (Hazan et al., sous presse dans Genomics).

A ce jour, trois gènes responsables de deux formes de PSF-X et d'une forme de PSF-AR ont été identifiés. Des mutations dans le gène, codant pour une molécule d'adhérence cellulaire spécifique des neurones, L1-CAM (pour L1 Cell Adhesion Molecule), et localisé en Xq28 (locus SPG1) causent une forme complexe de PSF (14) dans laquelle la spasticité est associée à un retard mental, alors que des mutations dans le gène PLP (pour Proteolipid Protein) situé en Xq21 (locus SPG2), codant pour une molécule constitutive de la couche de myéline, sont à l'origine de formes pures et complexes de PSF-X (15). Plus récemment des mutations dans un gène localisé en 16q24.3 (locus SPG7), qui code pour la paraplegin, une ATPase mitochondriale de la famille protéique des AAA (pour «ATPases Associated with diverse cellular Activities») (16) ont été associées à des formes complexes et pures de PSF-AR (17), suggérant que des altérations de la phosphorylation oxydative (OXPHOS) peuvent être à l'origine de PSF.

Ainsi, il reste aujourd'hui un grand besoin d'identifier et de caractériser le gène responsable de la forme la plus fréquente de PSF-AD. L'identification de ce gène devrait en particulier permettre, outre la possibilité d'un test de dépistage anténatal chez les familles concernées, de mieux comprendre certains des mécanismes moléculaires engendrant ces dégénérescences spécifiques de faisceaux nerveux de la moelle épinière, voire d'apporter des éléments de réponse quant à un traitement thérapeutique des malades.

Ceci est justement l'objet de la présente invention.

Après avoir délimité l'intervalle de localisation entre les marqueurs génétiques D2S352 et D2S2347 par l'étude des événements de recombinaison dans les familles

de PSF liées au locus SPG4, les inventeurs ont établi un contig de BACs recouvrant une distance physique évaluée à environ 1,5 Mb et entrepris une stratégie de clonage positionnel basée sur le séquençage de l'intervalle SPG4 afin d'identifier de façon exhaustive tous les gènes localisés dans la région candidate. L'analyse de la séquence des deux BACs, D (b336P14) et G (B763N4), a révélé la présence d'un gène composé de 17 exons, s'étendant sur une distance d'environ 100 kb, et présentant une homologie avec les gènes codant pour des protéines de la famille des AAA. La comparaison de la séquence de ce gène entre les individus sains et atteints des familles de PSF-AD a permis de mettre en évidence différentes mutations chez les patients.

L'invention a ainsi pour objet l'identification et la caractérisation du gène SPG4 (ou SPAST) codant pour un nouveau membre nucléaire de la famille des AAA, responsable de la forme la plus fréquente de PSF-AD.

Sous un premier aspect, la présente invention a pour objet un acide nucléique purifié ou isolé du gène SPG4, caractérisé en ce qu'il comprend au moins 15 nucléotides consécutifs, de préférence 20, 25, 30, 50, 100 ou 200 nucléotides consécutifs, d'une séquence choisie parmi le groupe comprenant :

- la séquence SEQ ID No. 1, séquence génomique du gène SPG4 humain ;
- les séquences nucléiques homologues ou variantes de l'acide nucléique de séquence SEQ ID No. 1 ;
- leur séquence complémentaire ; et
- la séquence de leur ARN correspondant.

La présente invention concerne, bien entendu, aussi bien les séquences ADN qu'ARN ainsi que les séquences qui s'hybrident avec elles, de même que les ADNs double brin correspondants.

Par acide nucléique, séquence nucléique ou d'acide nucléique, polynucléotide, oligonucléotide, séquence de polynucléotide, séquence nucléotidique, termes qui seront employés indifféremment dans la présente description, on entendra désigner aussi bien un ADN double brin, un ADN simple brin que des produits de transcription desdits ADNs, et/ou un fragment d'ARN, lesdits fragments naturels isolés, ou de synthèse, comportant ou non des nucléotides non naturels, désignant un enchaînement précis de nucléotides, modifiés ou non, permettant de définir un fragment ou une région d'un acide nucléique. On entend par fragment d'ADN et/ou d'ARN naturel isolé, ou de synthèse, comportant ou non des nucléotides non naturels,

un enchaînement précis de nucléotides, modifiés ou non, permettant de définir un fragment, un segment ou une région d'un acide nucléique.

Il doit être compris que la présente invention ne concerne pas les séquences nucléotidiques génomiques dans leur environnement chromosomique naturel, c'est-à-dire à l'état naturel. Il s'agit de séquences qui ont été isolées et/ou purifiées, c'est-à-dire qu'elles ont été prélevées directement ou indirectement, par exemple par copie, leur environnement ayant été au moins partiellement modifié.

Par «séquence nucléique homologue», on entendra désigner les séquences nucléiques présentant, par rapport à la séquence nucléique de référence certaines modifications comme en particulier une délétion, une troncation, un allongement, une fusion chimérique, et/ou une mutation, notamment ponctuelle, et dont la séquence nucléique présente au moins 80 %, de préférence 90 % ou 95 %, d'identité après alignement avec la séquence nucléique de référence. Il s'agit de préférence de séquences dont les séquences complémentaires sont susceptibles de s'hybrider spécifiquement avec une des séquences de l'invention. De préférence, les conditions d'hybridation spécifiques ou de forte stringence seront telles qu'elles assurent au moins 80 %, de préférence 90 % ou 95 % d'identité après alignement entre l'une des deux séquences et la séquence complémentaire de l'autre.

Une hybridation dans des conditions de forte stringence signifie que les conditions de température et de force ionique sont choisies de telle manière qu'elles permettent le maintien de l'hybridation entre deux fragments d'ADN complémentaires. A titre illustratif, des conditions de forte stringence de l'étape d'hybridation aux fins de définir les fragments polynucléotidiques décrits ci-dessus, sont avantageusement les suivantes :

L'hybridation ADN-ADN ou ADN-ARN est réalisée en deux étapes : (1) préhybridation à 42°C pendant 3 heures en tampon phosphate (20 mM, pH 7,5) contenant 5 x SSC (1 x SSC correspond à une solution 0,15 M NaCl + 0,015 M citrate de sodium), 50 % de formamide, 7 % de sodium dodécyl sulfate (SDS), 10 x Denhardt's, 5 % de dextran sulfate et 1 % d'ADN de sperme de saumon ; (2) hybridation proprement dite pendant 20 heures à une température dépendant de la taille de la sonde (i.e. : 42°C, pour une sonde de taille > 100 nucléotides) suivie de 2 lavages de 20 minutes à 20°C en 2 x SSC + 2 % SDS, 1 lavage de 20 minutes à 20°C en 0,1 x SSC + 0,1 % SDS. Le dernier lavage est pratiqué en 0,1 x SSC + 0,1 % SDS pendant 30 minutes à 60°C pour une sonde de taille > 100 nucléotides. Les conditions d'hybridation de forte stringence décrites ci-dessus pour un polynucléotide de taille

définie, seront adaptées par l'homme du métier pour des oligonucléotides de taille plus grande ou plus petite, selon l'enseignement de Sambrook et al., 1989.

Par «séquence nucléique variante» ou «acide nucléique variant» d'une séquence nucléique de référence, on entendra désigner l'ensemble des séquences nucléiques correspondant à des variants alléliques, c'est-à-dire des variations individuelles de la séquence nucléique de référence. Ces séquences mutées naturelles correspondent à des polymorphismes présents chez les mammifères, en particulier chez l'être humain et, notamment, à des polymorphismes pouvant conduire à la survenue et/ou au développement de pathologie.

Si les séquences selon l'invention concernent les séquences normales, elles concernent également les séquences mutées dans la mesure où elles comportent au moins une mutation ponctuelle et de préférence au plus 10 % de mutations par rapport à la séquence normale.

En particulier, les séquences nucléiques variantes comprendront toute séquence d'au moins 15 nucléotides consécutifs, de préférence 20, 25, 30, 50, 100 ou 200 nucléotides consécutifs, d'une séquence polymorphique de la séquence génomique du gène SPG4 humain de séquence SEQ ID No. 1, et dont la séquence d'acide nucléique présente par rapport à la séquence SEQ ID No. 1 au moins une mutation correspondant notamment à une troncation, délétion, substitution et/ou addition de résidu d'acide aminé. Dans le cas présent, les séquences nucléiques variantes présentant au moins une mutation seront ici liées aux pathologies de type PSF-AD liées au locus SPG4.

De préférence, la présente invention concerne des séquences nucléiques mutées dans lesquelles les mutations conduisent à une modification de la séquence d'acides aminés du polypeptide codé par la séquence normale.

On entendra également désigner par séquences nucléiques variantes tout ARN ou ADNc résultant d'une mutation d'un site d'épissage de la séquence nucléique génomique SEQ ID No. 1.

L'invention concerne de préférence un acide nucléique purifié ou isolé selon la présente invention, caractérisé en ce qu'il comprend une séquence choisie parmi le groupe suivant :

- la séquence SEQ ID No. 1 ;
- la séquence SEQ ID No. 2, séquence de l'ADNc codant pour la spastin humaine ;
- la séquence SEQ ID No. 72, séquence de l'ADNc incomplet codant pour la spastin murine représentée à la figure 5, ligne "Mouse" ;

- les séquences nucléiques homologues ou variantes des séquences SEQ ID No. 1, SEQ ID No. 2 ou SEQ ID No. 72 ;
- leur séquence complémentaire ; et
- la séquence de leur ARN correspondant.

5 De préférence, l'invention concerne un acide nucléique purifié ou isolé selon l'invention, caractérisé en ce qu'il comprend au moins une mutation dont la position et la nature sont identifiées dans le tableau 5.

Les amorces ou sondes, caractérisées en ce qu'elles comprennent une séquence d'un acide nucléique selon l'invention, font également partie de l'invention.

10 La présente invention concerne ainsi l'ensemble des amorces qui peuvent être déduites des séquences nucléotidiques de l'invention et qui peuvent permettre de mettre en évidence lesdites séquences nucléotidiques de l'invention, en particulier les séquences mutées, en utilisant notamment une méthode d'amplification telle que la méthode PCR, ou une méthode apparentée.

15 La présente invention concerne également l'ensemble des sondes qui peuvent être déduites des séquences nucléotidiques de l'invention, notamment des séquences capables de s'hybrider avec elles, et qui peuvent permettre de mettre en évidence lesdites séquences nucléotidiques, en particulier de discriminer les séquences normales des séquences mutées.

20 La présente invention concerne en particulier les sondes ou amorces de séquences choisies parmi les séquences SEQ ID No. 4 à SEQ ID No. 71.

L'invention concerne également l'utilisation d'une séquence d'acide nucléique selon l'invention comme sonde ou amorce, pour la détection, l'identification, le dosage ou l'amplification de séquence d'acide nucléique.

25 Selon l'invention, les polynucléotides pouvant être utilisés comme sonde ou comme amorce dans des procédés de détection, d'identification, de dosage ou d'amplification de séquence nucléique, présenteront une taille minimale de 15 bases, de préférence de 20 bases, ou mieux de 25 à 30 bases.

30 L'ensemble des sondes et amorces selon l'invention pourront être marquées directement ou indirectement par un composé radioactif ou non radioactif par des méthodes bien connues de l'homme du métier, afin d'obtenir un signal détectable et/ou quantifiable.

Les séquences de polynucléotides selon l'invention non marquées peuvent être utilisées directement comme sonde ou amorce.

Les séquences sont généralement marquées pour obtenir des séquences utilisables pour de nombreuses applications. Le marquage des amorces ou des sondes selon l'invention est réalisé par des éléments radioactifs ou par des molécules non radioactives.

5 Parmi les isotopes radioactifs utilisés, on peut citer le ^{32}P , le ^{33}P , le ^{35}S , le ^3H ou le ^{125}I . Les entités non radioactives sont sélectionnées parmi les ligands tels la biotine, l'avidine, la streptavidine, la dioxygénine, les haptènes, les colorants, les agents luminescents tels que les agents radioluminescents, chémiluminescents, bioluminescents, fluorescents, phosphorescents.

10 Les polynucléotides selon l'invention peuvent ainsi être utilisés comme amorce et/ou sonde dans des procédés mettant en oeuvre notamment la technique de PCR (réaction en chaîne à la polymérase) (Erich, 1989 ; Innis et al., 1990, et Rolfs et al., 1991). Cette technique nécessite le choix de paires d'amorces oligonucléotidiques encadrant le fragment qui doit être amplifié. On peut, par exemple, se référer à la
15 technique décrite dans le brevet américain U.S. N° 4,683,202. Les fragments amplifiés peuvent être identifiés, par exemple après une électrophorèse en gel d'agarose ou de polyacrylamide, ou après une technique chromatographique comme la filtration sur gel ou la chromatographie échangeuse d'ions, puis séquencés. La spécificité de l'amplification peut être contrôlée en utilisant comme amorce les séquences
20 nucléotidiques de polynucléotides de l'invention comme matrice, des plasmides contenant ces séquences ou encore les produits d'amplification dérivés. Les fragments nucléotidiques amplifiés peuvent être utilisés comme réactifs dans des réactions d'hybridation afin de mettre en évidence la présence, dans un échantillon biologique, d'un acide nucléique cible de séquence complémentaire à celle desdits fragments
25 nucléotidiques amplifiés.

L'invention vise également les acides nucléiques susceptibles d'être obtenus par amplification à l'aide d'amorces selon l'invention.

D'autres techniques d'amplification de l'acide nucléique cible peuvent être avantageusement employées comme alternative à la PCR (PCR-like) à l'aide de
30 couple d'amorces de séquences nucléotidiques selon l'invention. Par PCR-like on entendra désigner toutes les méthodes mettant en oeuvre des reproductions directes ou indirectes des séquences d'acides nucléiques, ou bien dans lesquelles les systèmes de marquage ont été amplifiés, ces techniques sont bien entendu connues, en général il s'agit de l'amplification de l'ADN par une polymérase ; lorsque
35 l'échantillon d'origine est un ARN il convient préalablement d'effectuer une transcription

reverse. Il existe actuellement de très nombreux procédés permettant cette amplification, comme par exemple la technique SDA (Strand Displacement Amplification) ou technique d'amplification à déplacement de brin (Walker et al., 1992), la technique TAS (Transcription-based Amplification System) décrite par Kwoh et al. en 5 1989, la technique 3SR (Self-Sustained Sequence Replication) décrite par Guatelli et al. en 1990, la technique NASBA (Nucleic Acid Sequence Based Amplification) décrite par Kievitis et al. en 1991, la technique TMA (Transcription Mediated Amplification), la technique LCR (Ligase Chain Reaction) décrite par Landegren et al. en 1988 et perfectionnée par Barany et al. en 1991, qui emploie une ligase thermostable, la 10 technique de RCR (Repair Chain Reaction) décrite par Segev en 1992, la technique CPR (Cycling Probe Reaction) décrite par Duck et al. en 1990, la technique d'amplification à la Q-béta-réplase décrite par Miele et al. en 1983 et perfectionnée notamment par Chu et al. en 1986 et Lizardi et al. en 1988, puis par Burg et al. ainsi que par Stone et al. en 1996.

15 Dans le cas où le polynucléotide cible à détecter est un ARNm, on utilisera avantageusement, préalablement à la mise en oeuvre d'une réaction d'amplification à l'aide des amorces selon l'invention ou à la mise en oeuvre d'un procédé de détection à l'aide des sondes de l'invention, une enzyme de type transcriptase reverse afin d'obtenir un ADNc à partir de l'ARNm contenu dans l'échantillon biologique. L'ADNc 20 obtenu servira alors de cible pour les amorces ou les sondes mises en oeuvre dans le procédé d'amplification ou de détection selon l'invention.

La technique d'hybridation de sondes peut être réalisée de manières diverses (Matthews et al., 1988). La méthode la plus générale consiste à immobiliser l'acide nucléique extrait des cellules de différents tissus ou de cellules en culture sur un 25 support (tels que la nitrocellulose, le nylon, le polystyrène) et à incubé, dans des conditions bien définies, l'acide nucléique cible immobilisé avec la sonde. Après l'hybridation, l'excès de sonde est éliminé et les molécules hybrides formées sont détectées par la méthode appropriée (mesure de la radioactivité, de la fluorescence ou de l'activité enzymatique liée à la sonde).

30 Selon un autre mode de mise en oeuvre des sondes nucléiques selon l'invention, ces dernières peuvent être utilisées comme sonde de capture. Dans ce cas, une sonde, dite « sonde de capture », est immobilisée sur un support et sert à capturer par hybridation spécifique l'acide nucléique cible obtenu à partir de l'échantillon biologique à tester et l'acide nucléique cible est ensuite détecté grâce à une seconde 35 sonde, dite « sonde de détection », marquée par un élément facilement détectable.

Les séquences de site accepteur ou donneur d'épissage identifiées au tableau 3 font également partie de la présente invention.

Sous un autre aspect, l'invention comprend une méthode pour le criblage de banques d'ADNc ou d'ADN génomique, ou pour le clonage d'ADNc ou génomique isolé codant pour la spastin, caractérisée en ce qu'elle met en œuvre une séquence nucléique selon l'invention.

Parmi ces méthodes, on peut citer notamment :

- le criblage de banques d'ADNc et le clonage des ADNc isolés (Sambrook et al., 1989 ; Suggs et al., 1981 ; Woo et al., 1979), à l'aide des séquences nucléiques selon l'invention ;
- le criblage de banques génomiques, par exemple de BACs (Chumakov et al., 1992 ; Chumakov et al., 1995) et éventuellement une analyse génétique en FISH (Cherif et al., 1990) à l'aide de séquences selon l'invention, permettant l'isolement et la localisation chromosomique, puis le séquençage complet du gène SPG4 codant pour la spastin.

En particulier, ces méthodes selon l'invention pourront être mises en œuvre pour l'identification et ainsi l'obtention de la séquence génomique ou de l'ADNc du gène SPG4 chez d'autres mammifères, notamment la souris.

Ces méthodes de criblage et/ou de clonage comprendront en particulier une étape d'hybridation d'un acide nucléique selon l'invention avec un acide nucléique contenu dans une banque génomique ou d'ADNc.

L'invention comprend aussi une méthode d'identification des séquences d'acide nucléique promotrices et/ou régulatrices de l'expression du gène SPG4 de séquence SEQ ID No. 1, caractérisée en ce qu'elle met en œuvre un acide nucléique selon l'invention.

Les outils informatiques à la disposition de l'homme du métier lui permettent aisément d'identifier à partir des séquences nucléiques génomiques selon l'invention les boîtes régulatrices promotrices nécessaires et suffisantes au contrôle de l'expression génique, notamment les boîtes TATA, CCAAT, GC, ainsi que les séquences régulatrices stimulatrices (« enhancer ») ou inhibitrices (« silencers ») qui contrôlent en CIS l'expression des gènes selon l'invention ; parmi ces séquences régulatrices, il convient de citer l'IRE, MRE, CRE.

L'invention concerne également des méthodes pour l'identification de mutations portées par le gène SPG4 humain caractérisées en ce qu'elles mettent en œuvre une

séquence nucléique selon l'invention, notamment de mutations responsables de la paraplégie spastique familiale autosomique dominante.

Ces méthodes d'identification de ces mutations comprendront en particulier les étapes suivantes : (i) isolement de l'ADN à partir de l'échantillon biologique à analyser, ou obtention d'un ADNc à partir de l'ARNm de l'échantillon biologique ; (ii) amplification spécifique de l'ADN cible susceptible de présenter une mutation à l'aide d'amorces selon l'invention ; (iii) analyse des produits d'amplification, notamment la taille et/ou la séquence des produits d'amplification, par rapport à une séquence de référence.

Par méthodes d'identification de mutation selon l'invention, on entend également désigner une méthode permettant d'obtenir l'acide nucléique sur lequel a été identifiée ladite mutation.

Font également partie de l'invention, les séquences promotrices et/ou régulatrices du gène SPG4 selon l'invention présentant des mutations susceptibles de modifier l'expression de la protéine correspondante.

Les acides nucléiques caractérisés en ce qu'ils sont susceptibles d'être obtenus par une des méthodes précédentes selon l'invention, ou les acides nucléiques capables de s'hybrider dans des conditions de forte stringence (homologie d'au moins 80 % entre une des deux séquences et la séquence complémentaire de l'autre) avec lesdits acides nucléiques, font partie de l'invention, notamment les acides nucléiques variants ou homologues, en particulier les séquences nucléiques de variants alléliques du gène SPG4 de séquence SEQ ID No. 1 ou de son ADNc de séquence SEQ ID No. 2, ainsi que les séquences génomiques des gènes homologues d'autres mammifères tels que la souris.

Dans la présente description, on entendra désigner par "Spg4" le gène de souris homologue au gène humain SPG4.

L'utilisation d'une séquence d'acide nucléique selon l'invention comme sonde ou amorce pour le criblage de banque génomique ou d'ADNc fait bien entendu partie de l'objet de la présente invention.

Sous un autre aspect, l'invention comprend un polypeptide purifié ou isolé codé par un acide nucléique selon l'invention.

Dans la présente description, on utilisera le terme polypeptide pour désigner également une protéine ou un peptide.

De préférence, la présente invention concerne un polypeptide, caractérisé en ce qu'il comprend une séquence d'acides aminés choisie parmi le groupe suivant :

- la séquence SEQ ID No. 3, correspondant à la spastin humaine codée par la séquence SEQ ID No. 2 de l'ADNc du gène SPG4 humain ;
- la séquence SEQ ID No. 73, correspondant à un fragment de la spastin murine codée par la séquence SEQ ID No. 72 de l'ADNc incomplet du gène Spg4 de souris, la
5 séquence SEQ ID No. 73 est représentée à la figure 4A, ligne "SPAST_MOUSE" ;
- les séquences de polypeptides homologues et variants du polypeptide de séquence SEQ ID No. 3 ou SEQ ID No. 73 ; et
- les séquences de leurs fragments d'au moins 8, 10, 15, 30 ou 50 acides aminés consécutifs.

10 De manière également préférée, l'invention a pour objet un polypeptide, caractérisé en ce qu'il comprend une séquence d'acides aminés choisie parmi le groupe suivant :

- la séquence SEQ ID No. 3 et la séquence SEQ ID No. 73, lesquelles séquences portant au moins une des mutations dont la nature et la localisation sont identifiées
15 au tableau 5 ci-après ; et
- les séquences de leurs fragments d'au moins 8, 10, 15, 30 ou 50 acides aminés consécutifs.

Il doit être compris que l'invention ne concerne pas les polypeptides sous forme naturelle, c'est-à-dire qu'ils ne sont pas pris dans leur environnement. En effet,
20 l'invention concerne les peptides obtenus par purification à partir de sources naturelles, ou bien obtenus par recombinaison génétique, ou encore par synthèse chimique et pouvant alors comporter des amino-acides non naturels. La production d'un polypeptide recombinant, qui peut être réalisée en utilisant l'une des séquences nucléotidiques selon l'invention est particulièrement avantageuse car elle permet
25 d'obtenir un niveau de pureté accrue du polypeptide désiré.

Par polypeptide homologue, on entendra désigner les polypeptides présentant certaines modifications par rapport au polypeptide de référence, comme en particulier une ou plusieurs délétions, troncations, un allongement, une fusion chimérique, et/ou une ou plusieurs substitutions, et dont la séquence d'acides aminés présente au moins
30 80 %, de préférence 90 % ou 95 %, d'identité après alignement avec la séquence d'acides aminés de référence.

Par « polypeptide variant » (ou variant protéique), on entendra désigner l'ensemble des polypeptides codés par les séquences nucléiques variantes telles que précédemment définies.

En particulier, les polypeptides variants comprendront tout polypeptide codé par la séquence génomique mutée du gène SPG4 de séquence SEQ ID N° 1, et dont la séquence d'acides aminés présente au moins une mutation correspondant notamment à une troncation, délétion, substitution et/ou addition de résidus d'acides aminés par rapport à la séquence SEQ ID No. 3. Dans le cas présent, les polypeptides variants
5 présentant au moins une mutation seront liés aux pathologies de type PSF-AD.

On entendra également désigner par polypeptide variant tout polypeptide résultant de mutation d'un site d'épissage dans la séquence nucléique génomique SEQ ID No. 1.

10 L'invention comprend également les vecteurs de clonage et/ou d'expression contenant une séquence d'acide nucléique selon l'invention.

Les vecteurs selon l'invention, caractérisés en ce qu'ils comportent les éléments permettant l'expression et/ou la sécrétion desdites séquences dans une cellule hôte, ou encore une séquence d'adressage cellulaire, font également partie de
15 l'invention.

Les vecteurs caractérisés en ce qu'ils comportent une séquence de promoteur et/ou de régulateur selon l'invention, font également partie de l'invention.

Lesdits vecteurs comporteront de préférence un promoteur, des signaux d'initiation et de terminaison de la traduction, ainsi que des régions appropriées de régulation de la transcription. Ils doivent pouvoir être maintenus de façon stable dans la
20 cellule et peuvent éventuellement posséder des signaux particuliers spécifiant la sécrétion de la protéine traduite.

Ces différents signaux de contrôle sont choisis en fonction de l'hôte cellulaire utilisé. A cet effet, les séquences d'acide nucléique selon l'invention peuvent être
25 insérées dans des vecteurs à répllication autonome au sein de l'hôte choisi, ou des vecteurs intégratifs de l'hôte choisi.

Parmi les systèmes à répllication autonome, on utilisera de préférence n fonction de la cellule hôte, des systèmes de type plasmidique ou viral, les vecteurs viraux pouvant notamment être des adénovirus (Perricaudet et al., 1992), des
30 rétrovirus, des lentivirus, des poxvirus ou des virus herpétiques (Epstein et al., 1992). L'homme du métier connaît les technologies utilisables pour chacun de ces systèmes.

Lorsque l'on souhaitera l'intégration de la séquence dans les chromosomes de la cellule hôte, on pourra utiliser par exemple des systèmes de type plasmidique ou viral ; de tels virus seront, par exemple, les rétrovirus (Temin, 1986), ou les AAV
35 (Carter, 1993).

Parmi les vecteurs non viraux, on préfère les polynucléotides nus tels que l'ADN nu ou l'ARN nu selon la technique développée par la société VICAL, les chromosomes artificiels de levure (YAC, yeast artificial chromosome) pour l'expression dans la levure, les chromosomes artificiels de souris (MAC, mouse artificial chromosome) pour
 5 l'expression dans les cellules murines et de manière préférée les chromosomes artificiels d'homme (HAC, human artificial chromosome) pour l'expression dans les cellules humaines.

De tels vecteurs seront préparés selon les méthodes couramment utilisées par l'homme du métier, et les clones en résultant peuvent être introduits dans un hôte
 10 approprié par des méthodes standard, telles que par exemple la lipofection, l'électroporation, le choc thermique.

L'invention comprend en outre les cellules hôtes, notamment les cellules eucaryotes et procaryotes, transformées par les vecteurs selon l'invention ainsi que les animaux transgéniques, excepté l'Homme, comprenant une desdites cellules
 15 transformées selon l'invention.

Parmi les cellules utilisables à ces fins, on peut citer bien entendu les cellules bactériennes (Olins et Lee, 1993), mais également les cellules de levure (Buckholz, 1993), de même que les cellules animales, en particulier les cultures de cellules de mammifères (Edwards et Aruffo, 1993), et notamment les cellules d'ovaire de hamster
 20 chinois (CHO), mais également les cellules d'insectes dans lesquelles on peut utiliser des procédés mettant en œuvre des baculovirus par exemple (Luckow, 1993). Un hôte cellulaire préféré pour l'expression des protéines de l'invention est constitué par les cellules CHO.

Parmi les mammifères selon l'invention, on préférera des animaux tels que les
 25 souris, les rats ou les lapins, exprimant un polypeptide selon l'invention.

Parmi les mammifères selon l'invention, on préférera également ceux comprenant une cellule transformée caractérisée en ce que la séquence de l'un au moins des deux allèles du gène SPG4 contient une au moins des mutations dont la position et la nature sont identifiées dans le tableau 5 ou identifiées par une méthode
 30 selon la présente invention.

Parmi les mammifères selon l'invention, on préférera également des animaux tels que les souris, les rats ou les lapins, caractérisés en ce que le gène codant pour la spastin selon l'invention, n'est pas fonctionnel ou est invalidé.

Parmi les modèles animaux plus particulièrement intéressants ici, on trouve
 35 notamment :

- les animaux transgéniques présentant au moins dans une de leurs deux séquences alléliques du gène SPG4, une au moins des mutations dont la position et la nature sont identifiées au tableau 5 ou identifiées par une méthode selon la présente invention. Ces animaux transgéniques sont obtenus par exemple par recombinaison
5 homologue sur cellules souches embryonnaires, transfert de ces cellules souches à des embryons, sélection des chimères affectées au niveau des lignées reproductrices, et croissance desdites chimères ;
- les animaux (de préférence souris) transgéniques surexprimant le gène SPG4 dans lequel pourra être introduite une desdites mutations selon l'invention. Les souris sont
10 obtenues par exemple par transfection de copie de ce gène sous contrôle d'un promoteur fort de nature ubiquitaire, ou sélectif d'un type de tissu, ou après transcription virale ;
- les animaux (de préférence souris) transgéniques rendus déficients pour le gène SPG4 selon l'invention, par inactivation à l'aide du système LOXP/CRE recombinase
15 (Rohlmann et al., 1996) ou de tout autre système d'inactivation de l'expression de ce gène.

Les cellules et mammifères selon l'invention sont utilisables dans une méthode de production d'un polypeptide selon l'invention, comme décrit ci-dessous, et peuvent également servir à titre de modèle d'analyse et pour le criblage de banque d'ADN
20 (génomique ou d'ADNc).

Les cellules ou mammifères transformés tels que décrits précédemment peuvent être ainsi utilisés à titre de modèles afin d'étudier les interactions entre les polypeptides selon l'invention, et les composés chimiques ou protéiques, impliqués directement ou indirectement dans les activités des polypeptides selon l'invention, ceci
25 afin d'étudier les différents mécanismes et interactions mis en jeu.

Surtout, ils peuvent être utilisés pour la sélection de produits interagissant avec les polypeptides selon l'invention, notamment la spastin humaine de séquence SEQ ID No. 3 ou ses variants selon l'invention, à titre de cofacteur, ou d'inhibiteur, notamment compétitif, ou encore ayant une activité agoniste ou antagoniste de l'activité des
30 polypeptides selon l'invention. De préférence, on utilisera lesdites cellules transformées ou animaux transgéniques à titre de modèle permettant, notamment, la sélection de produits permettant de lutter contre la pathologie liée au gène SPG4 mentionnée ci-dessus.

L'invention concerne également l'utilisation de cellule, de mammifère ou de
35 polypeptide selon l'invention pour le criblage de composé chimique ou biochimique

pouvant interagir directement ou indirectement avec les polypeptides selon l'invention, et/ou capable de moduler l'expression ou l'activité de ces polypeptides.

L'invention concerne également l'utilisation d'une séquence d'acide nucléique selon l'invention pour la synthèse de polypeptides recombinants.

5 La méthode de production d'un polypeptide de l'invention sous forme recombinante est elle-même comprise dans la présente invention, et se caractérise en ce que l'on cultive les cellules transformées, notamment les cellules ou mammifères de la présente invention, dans des conditions permettant l'expression d'un polypeptide recombinant codé par une séquence d'acide nucléique selon l'invention, et que l'on
10 récupère ledit polypeptide recombinant.

Les polypeptides recombinants, caractérisés en ce qu'ils sont susceptibles d'être obtenus par ladite méthode de production, font également partie de l'invention.

Les polypeptides recombinants obtenus comme indiqué ci-dessus, peuvent aussi bien se présenter sous forme glycosylée que non glycosylée et peuvent
15 présenter ou non la structure tertiaire naturelle.

Ces polypeptides peuvent être produits à partir des séquences d'acide nucléique définies ci-dessus, selon les techniques de production de polypeptides recombinants connues de l'homme du métier. Dans ce cas, la séquence d'acide nucléique utilisée est placée sous le contrôle de signaux permettant son expression
20 dans un hôte cellulaire.

Un système efficace de production d'un polypeptide recombinant nécessite de disposer d'un vecteur et d'une cellule hôte selon l'invention.

Ces cellules peuvent être obtenues par l'introduction dans des cellules hôtes d'une séquence nucléotidique insérée dans un vecteur tel que défini ci-dessus, puis la
25 mise en culture desdites cellules dans des conditions permettant la réplication et/ou l'expression de la séquence nucléotidique transfectée.

Les procédés de purification de polypeptide recombinant utilisés sont connus de l'homme du métier. Le polypeptide recombinant peut être purifié à partir de lysats et extraits cellulaires, du surnageant du milieu de culture, par des méthodes utilisées
30 individuellement ou en combinaison, telles que le fractionnement, les méthodes de chromatographie, les techniques d'immunoaffinité à l'aide d'anticorps monoclonaux ou polyclonaux spécifiques, etc..

Les polypeptides selon la présente invention peuvent être obtenus par synthèse chimique et ce en utilisant l'une des nombreuses synthèses peptidiques connues, par
35 exemple les techniques mettant en oeuvre des phases solides ou des techniques

utilisant des phases solides partielles, par condensation de fragments ou par une synthèse en solution classique.

La technique de synthèse en phase solide est bien connue de l'homme du métier. Voir notamment Stewart et al. (1984) et Bodansky (1984).

5 Les polypeptides obtenus par synthèse chimique et pouvant comporter des acides aminés non naturels correspondants sont également compris dans l'invention.

Les anticorps mono- ou polyclonaux ou leurs fragments, anticorps chimériques ou immunoconjugués, caractérisés en ce qu'ils sont capables de reconnaître spécifiquement un polypeptide selon l'invention, font partie de l'invention.

10 Des anticorps polyclonaux spécifiques peuvent être obtenus à partir d'un sérum d'un animal immunisé contre les polypeptides selon l'invention, notamment produit par recombinaison génétique ou par synthèse peptidique, selon les modes opératoires usuels.

On notera notamment l'intérêt d'anticorps reconnaissant de façon spécifique 15 certains polypeptides, variants, ou leurs fragments immunogènes, selon l'invention.

Les anticorps monoclonaux spécifiques peuvent être obtenus selon la méthode classique de culture d'hybridomes décrite par Köhler et Milstein, 1975.

Les anticorps selon l'invention sont, par exemple, des anticorps chimériques, des anticorps humanisés, des fragments Fab ou F(ab')₂. Ils peuvent également se 20 présenter sous forme d'immunoconjugués ou d'anticorps marqués afin d'obtenir un signal détectable et/ou quantifiable.

L'invention concerne également des méthodes pour la détection et/ou la purification d'un polypeptide selon l'invention, caractérisées en ce qu'elles mettent en œuvre un anticorps selon l'invention.

25 L'invention comprend en outre des polypeptides purifiés, caractérisés en ce qu'ils sont obtenus par une méthode selon l'invention.

Par ailleurs, outre leur utilisation pour la purification des polypeptides, les anticorps de l'invention, en particulier les anticorps monoclonaux, peuvent également être utilisés pour la détection de ces polypeptides dans un échantillon biologique.

30 Ils constituent ainsi un moyen d'analyse immunocytochimique ou immunohistochimique de l'expression des polypeptides selon l'invention, notamment le polypeptide de séquence SEQ ID No. 3 ou l'un de ses variants, sur des coupes de tissus spécifiques, par exemple par immunofluorescence, marquage à l'or, immunoconjugués enzymatiques.

Ils pourront permettre notamment de mettre en évidence une expression anormale de ces polypeptides dans les tissus ou prélèvements biologiques, ce qui les rend utiles pour le suivi de l'évolution de la maladie et le diagnostic moléculaire.

5 Plus généralement, les anticorps de l'invention peuvent être avantageusement mis en œuvre dans toute situation où l'expression d'un polypeptide selon l'invention, normal ou muté, doit être observée.

Font également partie de l'invention, les méthodes de détermination d'une variabilité allélique, d'une mutation, d'une délétion, d'une perte d'hétérozygotie ou de toute anomalie génétique du gène SPG4 selon l'invention, caractérisées en ce qu'elles
10 mettent en œuvre une séquence d'acide nucléique ou un anticorps selon l'invention.

La présente invention comprend ainsi une méthode de diagnostic génotypique de la pathologie associée au gène SPG4, caractérisée en ce que l'on met en œuvre une séquence d'acide nucléique selon l'invention.

De préférence, l'invention concerne une méthode de diagnostic génotypique de
15 la maladie associée à la présence d'au moins une mutation sur une séquence du gène SPG4 à partir d'un prélèvement biologique d'un patient, caractérisé en ce qu'il comporte les étapes suivantes :

- a) le cas échéant, isolement de l'ADN génomique à partir de l'échantillon biologique à analyser, ou obtention d'ADNc à partir de l'ARN de l'échantillon biologique ;
- 20 b) amplification spécifique de ladite séquence d'ADN du gène SPG4 susceptible de contenir une mutation à l'aide d'amorces selon l'invention ;
- c) analyse des produits d'amplification obtenus et comparaison de leur séquence avec la séquence normale correspondante du gène SPG4.

L'invention comprend également une méthode de diagnostic de la maladie
25 associée à une expression anormale d'un polypeptide codé par le gène SPG4, notamment le polypeptide de séquence SEQ ID No. 3, caractérisée en ce que l'on met en contact un ou des anticorps selon l'invention avec le matériel biologique à tester, dans des conditions permettant la formation éventuelle de complexes immunologiques spécifiques entre ledit polypeptide et le ou lesdits anticorps, et en ce que l'on détecte
30 et/ou quantifie les complexes immunologiques éventuellement formés.

Ces méthodes visent par exemple les méthodes de diagnostic de la PSF-AD associée à la présence de mutation dans le gène SPG4 selon l'invention, notamment anténatal, en déterminant à partir d'un prélèvement biologique du patient la présence
35 de mutations dans au moins une des séquences décrites précédemment. Les

séquences d'acides nucléiques analysées pourront aussi bien être de l'ADN génomique, de l'ADNc, ou de l'ARNm.

Des acides nucléiques ou anticorps basés sur la présente invention pourront également être utilisés pour permettre un diagnostic positif chez un malade ou un diagnostic pré-symptomatique chez un sujet à risque, notamment avec antécédent familial.

Les méthodes permettant de mettre en évidence une mutation dans un gène par rapport au gène sauvage sont, bien entendu, très nombreuses. On peut essentiellement les diviser en deux grandes catégories. Le premier type de méthode est celui dans lequel la présence d'une mutation est détectée par comparaison de la séquence mutée avec la séquence correspondante sauvage, et le second type est celui dans lequel la présence de la mutation est détectée de façon indirecte, par exemple par évidence de mésappariements dus à la présence de la mutation.

Ces méthodes peuvent mettre en oeuvre les sondes et amorces de la présente invention décrites. Il s'agit généralement de séquences nucléiques d'hybridation purifiées comprenant au moins 15 nucléotides, de préférence 20, 25 ou 30 nucléotides, caractérisées en ce qu'elles peuvent s'hybrider spécifiquement avec une séquence nucléique selon l'invention.

De préférence, les conditions d'hybridation spécifiques sont telles que celles définies précédemment ou dans les exemples. La longueur de ces séquences nucléiques d'hybridation peut varier de 15, 20 ou 30 à 200 nucléotides, particulièrement de 20 à 50 nucléotides.

Parmi les méthodes de détermination d'une variabilité allélique, d'une mutation, d'une délétion, d'une perte d'hétérozygotie ou d'une anomalie génétique, on préfère les méthodes comprenant au moins une étape d'amplification dite par PCR (réaction en chaîne par la polymérase) ou par PCR-like de la séquence cible selon l'invention susceptible de présenter une anomalie à l'aide de couple d'amorces de séquences nucléotidiques selon l'invention. Les produits amplifiés pourront être traités à l'aide d'enzyme de restriction approprié avant de procéder à la détection ou au dosage du produit ciblé.

Les mutations du gène SPG4 selon l'invention, peuvent être responsables de différentes modifications de son produit de traduction, modifications utilisables pour une approche diagnostique. En effet, les modifications d'antigénicité liées à ces mutations peuvent permettre la mise au point d'anticorps spécifiques. La discrimination du produit de gène muté peut être réalisée par ces méthodes. Toutes ces modifications

peuvent être utilisées dans une approche diagnostique grâce à plusieurs méthodes bien connues basées sur l'utilisation d'anticorps mono- ou polyclonaux reconnaissant le polypeptide normal ou des variants mutés, comme par exemple par RIA ou par ELISA.

5 Sous un autre aspect, l'invention comprend une méthode de sélection d'un composé chimique ou biochimique capable de prévenir et/ou de traiter la PSF-AD associée au gène SPG4, caractérisée en ce que l'on met en œuvre une séquence d'acide nucléique selon l'invention, un polypeptide selon l'invention, un vecteur selon l'invention, une cellule selon l'invention, un mammifère selon l'invention ou un anticorps
10 selon l'invention.

Sont également comprises dans l'invention, les méthodes de sélection de composés chimiques ou biochimiques capables d'interagir directement ou indirectement avec des polypeptides selon l'invention ou avec les acides nucléiques selon l'invention, et/ou permettant de moduler l'expression ou l'activité de ces
15 polypeptides, caractérisées en ce qu'elles comprennent la mise en contact d'un polypeptide selon l'invention, d'une cellule transformée selon l'invention, ou d'un mammifère selon l'invention, avec un composé candidat et, la détection d'une modification de l'activité dudit polypeptide.

Par exemple, mais sans s'y limiter, on peut citer une méthode d'identification de
20 molécules capables d'interagir avec un polypeptide selon l'invention en utilisant un système de double hybride bactérien ou levure tel que le Matchmaker Two Hybrid System 2, selon les instructions du manuel accompagnant le Matchmaker Two Hybrid System 2 (Catalogue N° K1604-1, Clontech).

Les acides nucléiques codant pour des protéines interagissant avec les
25 séquences promotrices et/ou régulatrices du gène SPG4 selon l'invention, peuvent être criblés et/ou sélectionnés en utilisant un système de simple hybride tel que celui décrit dans le manuel accompagnant le kit Matchmaker One-Hybrid System de Clontech (Catalog N° K1603-).

Sous un autre aspect, l'invention comprend l'utilisation d'acide nucléique ou de
30 polypeptide selon l'invention, d'un vecteur selon l'invention, d'une cellule selon l'invention, ou d'un mammifère selon l'invention, pour l'étude de l'expression ou de l'activité du gène SPG4.

D'autres caractéristiques et avantages de l'invention apparaissent dans la suite de la description avec les exemples et les figures dont les légendes sont représentées ci-après.

5 LEGENDES DES FIGURES

FIGURES 1A, 1B et 1C : Carte physique de l'intervalle SPG4 et organisation génomique de SPG4.

FIGURE 1A : La région candidate de 1,5 Mb est délimitée par les marqueurs génétiques D2S352 et D2S2347 indiqués en caractères gras. La position des marqueurs polymorphes et autres STSs est indiquée en caractères standards alors que la position des ESTs est indiquée en italique. Les clones de BAC constituant la carte de préséquençage sont représentés par des rectangles dont le nom figure au-dessus et la taille précise du clone au-dessous si celle-ci a pu être déterminée. Le nom des BACs A, B, C... est suivi d'une parenthèse contenant le nom du clone précédé d'un « b » si le clone est issu de la banque de BACs CITB_978_SKB ou d'un « B » si celui-ci provient de la banque RPCI-11.

FIGURE 1B : Représentation schématique du gène SPG4 qui chevauche les BACs D (b336P14) et G (B563N4). Les exons sont figurés comme des rectangles noirs avec leur nom au-dessus.

FIGURE 1C : Les cinq mutations identifiées dans sept familles de PSF-AD liées au locus SPG4 sont positionnées dans les exons 7, 11, 13 et dans le site accepteur d'épissage de l'intron 15.

FIGURE 2 : Séquence nucléique et protéique de l'ADNc SPG4 et de la spastin.

Les 17 barres verticales avec un nombre situé au-dessous représentent les jonctions entre les différents exons. Le codon ATG initiateur est localisé en position nt 126-128 et le codon STOP de terminaison en position nt 1974-1976. Cinq des mutations identifiées à ce jour, dont la perte de l'exon 16, sont indiquées en italique (nt 1210, nt 1468, nt 1520, nt 1620 et pour la perte de l'exon 16 : nt 1813-1853). Le site de polyadénylation est en italique et souligné. Le signal de localisation nucléaire (NLS) putatif, RGKKK, ainsi que les trois domaines conservés prédits par l'analyse dans la base de données ProDom sont respectivement localisés aux positions aa7-11 (NLS), aa342-409 (domaine 92), aa411-509 (domaine 179) et aa512-599 (domaine 6226). Les quatre motifs prédits par la comparaison de séquence dans la base de données Prosite sont : deux motifs «leucine zippers» aux positions aa50-78 et aa508-529, le site de fixation de l'ATP (ou motif A de Walker) aux positions aa382-389 et le domaine

de dimérisation « helix-loop-helix » aux positions aa478-486. Les motifs A et B de Walker, «GPPGNGKT» et «IIFIDE», ainsi que le consensus minimal des AAA sont soulignés.

FIGURES 3A, 3B et 3C : Caractérisation d'une mutation d'un site d'épissage chez les
5 individus atteints de trois familles de PSF-AD liée au locus SPG4.

FIGURE 3A : Amplification par PCR du fragment IV de l'ADNc SPG4 à partir d'ADNc de lymphoblastes : puits M, marqueur de taille VII (Boehringer) ; puits 1, membre non atteint de la famille 2992 ; puits 2, patient de la famille 2992 ; puits 3, membre non atteint de la famille 5330 ; puits 4, patient de la famille 5330 ; puits 5,
10 patient de la famille 5226 ; puits 6, témoin négatif (ADN génomique humain).

FIGURE 3B : Graphe de séquence de la mutation du site accepteur d'épissage de l'intron 15.

Séquence génomique de l'individu contrôle en haut et d'un patient de la famille 2992 en bas. L'astérisque à la position nt 1813-4 indique un polymorphisme A->C qui
15 touche un nucléotide non conservé du site accepteur d'épissage de l'intron 15 chez le patient.

FIGURES 4A et 4B : Les homologies de la spastin.

Les résidus conservés et identiques sont respectivement surlignés en bleu et
jaune.

20 FIGURE 4A : Alignement multiple créé par CLUSTAL W de huit protéines issues de divers organismes et présentant une forte homologie de séquence avec les spastin humaine et murine (SEQ ID No. 73).

FIGURE 4B : Alignement par CLUSTAL W des métalloprotéases de levure AFG3, RCA1 et YME1, et des paraplegin et spastin humaines.

25 FIGURE 5 : Alignement par BLASTN des séquences nucléiques de l'ADNc SPG4 et de son orthologue de souris Spg4 (SEQ ID No. 72). Le site de polyadénylation de l'ADNc murin est souligné et en italique. Le codon STOP est localisé en position nt 1515-1517 dans l'ADNc murin et en position nt 1974-1976 dans l'ADNc humain.

FIGURES 6A, 6B et 6C : Analyse par PCR de l'expression de SPG4 et de son
30 orthologue murin Spg4.

FIGURE 6A : Collection d'ADNc provenant de multiples tissus de souris.

Puits M, marqueur de taille V (Boehringer) ; puits 1, cœur ; puits 2, cerveau ; puits 3, rate ; puits 4, poumon ; puits 5, foie ; puits 6, muscle squelettique ; puits 7, rein ; puits 8, testicule ; puits 9, embryon de 7 jours E7 ; puits 10, embryon de 11 jours

E11 ; puits 11, embryon de 15 jours E15 ; puits 12, embryon de 17 jours E17 ; puits 13, témoin négatif (ADN génomique de souris).

FIGURE 6B : Collection d'ADNc provenant de multiples tissus humains.

Puits M, marqueur de taille VII (Boehringer) ; puits 1, cerveau ; puits 2, cœur ;
 5 puits 3, rein ; puits 4, foie ; puits 5, poumon ; puits 6, pancréas ; puits 7, placenta ; puits 8, muscle squelettique ; puits 9, témoin négatif (ADN génomique humain) ; puits 10, témoin négatif (pas d'ADN).

FIGURE 6C : Collection d'ADNc provenant de multiples tissus de fœtus humain.

10 Puits M, marqueur de taille VII (Boehringer) ; puits 1, cerveau ; puits 2, cœur ; puits 3, rein ; puits 4, foie ; puits 5, poumon ; puits 6, muscle squelettique ; puits 7, rate ; puits 8, thymus ; puits 9, témoin négatif (ADN génomique humain) ; puits 10, témoin négatif (pas d'ADN).

15 **EXEMPLES**

Exemple 1 : Matériels et méthodes

1) Sous-clonage et séquençage de la région candidate

Douze BACs provenant de deux banques génomiques humaines, CITB_978_SKB (commercialisées par Research Genetics) et RPCI-11 (18), et
 20 recouvrant l'intervalle SPG4 ont été sélectionnés pour être séquencés (Hazan et al., sous presse dans Genomics). 40 µg de l'ADN de chaque BAC a été digéré partiellement par l'enzyme de restriction CviJI (CHIMERx) et séparé par électrophorèse sur gel d'agarose 0,4 % LMP (FMC). Des fractions d'ADN dont les tailles varient autour de 3, 5 et 10 kb ont été éluées avec la β-agarase (Biolabs) et liguées à un vecteur
 25 plasmide pBAM3 préalablement digéré par SmaI et déphosphorylé dans un rapport de 1Xinsert pour 5Xvecteur. Des bactéries E. coli DH10B électrocompétentes (GIBCO-BRL) ont été transformées par électroporation avec les différentes ligations. Environ 1000 à 1500 sous-clones par BAC (8 à 10 génomes équivalents) composés de 20 % de clones avec inserts à 10 kb, 40 % de clones avec inserts à 5 kb et 40 % de clones
 30 avec inserts à 3 kb ont été isolés. Les extrémités des inserts de ces clones ont été séquencées sur un séquenceur automatique LICOR 4200. Pour chaque BAC, les séquences furent assemblées en un squelette constitué de plusieurs contigs à l'aide des logiciels Phred et Phrap. Les trous entre chaque contig furent séquencés avec des dideoxynucléotides marqués sur un séquenceur ABI 377 (PE-Applied Biosystems). Les
 35 exons contenus dans ces contigs de séquence ont été prédits par les programmes

informatiques GRAIL II, GENSCAN, FGENEH et Genie. Les séquences furent également comparées dans les bases de données nucléiques et protéiques de l'EMBL et de GenBank avec les programmes BLASTN et BLASTX. La détermination des séquences promotrices fut réalisée par les programmes informatiques TSSG et TSSW.

- 5 Les résultats de toutes ces analyses de séquences furent visualisés par le programme d'annotation de séquence Genotator.

2) Clonage de l'ADNc

- 10 L'ADNc du gène SPG4 a été isolé par des expériences de RACE-PCR en 5' et 3' sur des ARN polyA+ de cerveau fœtal, cerveau adulte et foie adulte à l'aide du kit d'amplification d'ADNc Marathon (Clontech), selon les instructions du fournisseur. Une première PCR, suivie d'une PCR interne furent effectuées avec différents couples d'amorces dont les séquences sont indiquées sur le tableau 1 ci-après :

Tableau 1
Amorces utilisées pour les RACE-PCR et les amplifications d'ADNc

Amorce	Séquence (5'-3')	Position en 5' couple/ PCR Taille du produit		
SPA_5RACE5	CGGAGCTCCTCTTGGCTGCCATG	nt 405		
SPA_5RACE6	AGAAGCGCTGGCAGAGCCACACGAAG	nt 372		
SPA_5RACE7	AAGGCGACCAAACGCAGCAGCGCGAAG	nt 331		
SPA_3RACE1	AGGAGCAAGCTGTGGAATGGTATAAG	nt 550		
SPA_3RACE2	TGGTTATGGCCAAGGACCGCTTACAAC	nt 689		
SPA_3RACE3	CAAACGGACGTCTATAATGACAGTAC	nt 747		
SPA_3RACE4	TTAGGAATGTGGACAGCAACCTTGC	nt 1075		
SPA_3RACE5	CTTCTCTGAGGCCTGAGTTGTTTAC	nt 1207		
SPA_3RACE6	TGCTAGAATGACTGATGGATACTCAGG	nt 1736		
SPA_3RACE7	AGATGCAGCACTGGGTCCTATCCG	nt 1787		
SPA_3RACE8	ATGAACGTCATCGGCTACAGAAACAG	nt 2037		
SPA_Db	TAGCAGTGGCTGCCGCCGT	nt 45	b+m	655 pb
SPA_Dm	AAGCGGTCCTTGGCCATAAC	nt 700		
SPA_Dc	GGCGGCAGTGAGAGCTGTG	nt 106	c+n	543 pb
SPA_Dn	CTAGCTCTTTCACACTGTTC	nt 649		
SPA_Ad	AACAGGCCTTCGAGTACATC	nt 487	d+n	746 pb
SPA_Am	CTGTGAACAACCTCAGGCCTC	nt 1233		
SPA_Ac	ATGAGAAAGCAGGACAGAAG	nt 532		
SPA_An	TGCCAAGTCTTGACCAGC	nt 1175		
SPA_Ba	CTACAACTGCTACTCGTAAG	nt 1036	a+m	763 pb
SPA_Bm	CAGTGCTGCATCTTTTGCC	nt 1799		
SPA_Bb	TAGGAATGTGGACAGCAACC	nt 1076		
SPA_Bn	AAAGCTGTTAGGTCACCTTC	nt 1780		
SPA_Ca	TGGAGATGACAGAGTACTTG	nt 1550	a+m	766 pb
SPA_Cm	CTGGAATACTTTCATCTGC	nt 2316		
SPA_Cb	ATGAGGCTGTTCTCAGGCG	nt 1603		

Les produits de RACE-PCR ont été clonés avec le kit TA-cloning (Invitrogen) et les clones correspondants ont été séquencés sur un ABI 377 (PE-Applied Biosystems). La séquence du transcrit SPG4 a été vérifiée par le séquençage de produits de PCR amplifiés à partir d'une population d'ADNc provenant des lymphoblastes de 6 individus sains.

3) Détection des mutations

Les ARNs totaux ont été extraits de lignées de lymphoblastes d'un individu atteint par famille étudiée et de 6 individus témoins à l'aide du kit RNA PLUSR (bioprobe System). La synthèse de l'ADNc a été réalisée sur 500 ng à 1 µg d'ARN avec 100 pmoles d'amorces hexamères aléatoires (Pharmacia) et 200 unités de reverse transcriptase Superscript II (Gibco BRL) dans des conditions standards. Quatre amplifications par PCR, générant des fragments chevauchants qui recouvrent la totalité de la phase ouverte de lecture de SPG4, ont été réalisées sur les ADNc des patients et contrôles. Le fragment I fut amplifié avec les amorces SPA_Db/SPA_Dm, puis en PCR interne avec les amorces SPA_Dc/SPA_Dn. Les fragments II, III, et IV ont été respectivement amplifiés avec les amorces SPA_Ad/SPA_Am, SPA_Ba/SPA_Bm et SPA_Ca/SPA_Cm (cf. les séquences de ces amorces sur le tableau 1). Chaque amplification a été effectuée dans un volume total de 50 µl contenant 4 µl d'ADNc (~1/7^{ème} de la prép.), 20 pmoles de chaque amorce, 200 µM de dNTPs, 50 mM de KCl, 10 mM de Tris pH9, 1,5 mM MgCl₂, 0,1 % de triton X-100, 0,01 % de gélatine et 2,5 unités de Taq polymérase (Cetus-PE). Les réactions de PCR ont été réalisées selon le procédé du « hot start » : la Taq polymérase est ajoutée à 92°C après une première étape de dénaturation de 5 min à 94°C. Les échantillons sont par la suite soumis à 35 cycles de dénaturation (94°C pendant 40 sec), d'hybridation (55°C pendant 50 sec, à l'exception du fragment I : 58°C pendant 50 sec) et d'élongation (72°C pendant 1 min), suivis d'une dernière étape d'élongation (5 min à 72°C). Les produits de PCR sont séquencés sur un séquenceur automatique ABI 377 (PE-Applied Biosystems) avec les amorces SPA_Dc/SPA_Dn, SPA_Ac/SPA_An, SPA_Bb/SPA_Bn et SPA_Cb/SPA_Cm pour les fragments I, II, III et IV respectivement.

Les mutations furent également recherchées ou confirmées par séquençage des 17 exons prédits du gène SPG4 chez les patients et contrôles. Chaque exon fut amplifié avec le couple correspondant d'amorces « a+m » (cf. tableau 2 ci-après), à l'exception de l'exon 1 (gSPAex1c/gSPAex1m), et des exons 10, 11 et 12 qui furent co-amplifiés avec les couples d'amorces gSPAex10a/gSPAex12m et gSPAex11a/gSPAex12m.

Tableau 2
Amorces de PCR pour l'amplification et le séquençage des exons

Exon	Taille du produit	Programme de PCR	Amorce	Séquence (5'-3')
1	1048 pb	0	gSPAex1c	GTGAGCCGAACTGCACATTG
			gSPAex1m	CAAAGTCGACAGCTACAGTGC
			gSPAex1d	GGAAGTGTAGTTGAGTGGGA
			gSPAex1n	AGATGAGGCTCCGACCTAC
2	624 pb	3	gSPAex2a	AATGCCACACTTGTAATCTC
			gSPAex2m	TGTGAATATATCATAATTTGGG
			gSPAex2b	TACAGCAGTTCTCATGATG
3	812 pb	1	gSPAex3a	GACCAAATTGGTGCATGCATG
			gSPAex3m	ACATTTCCAATACATCCCAC
4	379 pb	3	gSPAex4a	ATTTGTCATTTACATGCAC
			gSPAex4m	TTAGAATGACTATACCTGAC
			gSPAex4n	TCAGGTAAAGTAAGACTC
5	830 pb	4	gSPAex5a	TTCTATCTACCTAGTGAC
			gSPAex5m	TTTTATAGCAAGTTGCCCTG
			gSPAex5b	CCTATGAAGATCCTGGTAC
6	484 pb	3	gSPAex6a	TGTCATGATTCTAACAAGGG
			gSPAex6m	TCTATTTCACTCCTGACATG
7	420 pb	2	gSPAex7a	GTCATAGGGCTTAGGCTTC
			gSPAex7m	ATCATACTACCCACTTTTCC
8	647 pb	3	gSPAex8a	TGTTTGGGAAGATGCTACTG
			gSPAex8m	CTACTGAAGATAACGTACATG
9	1268 pb	1	gSPAex9a	CATTGATTGCCATGTATTGG
			gSPAex9m	AGAAGGCCAGAAATACTCAG
			gSPAex9b	GTACTIONAATCGGTAAATATGG
10]	1061 pb	4	gSPAex10a	CTCAAGTCTTAGGAATGCAG
11			gSPAex10b	GCACTTAACCAGGCTGTATG
12]	551 pb	3	gSPAex11a	CTCAGATGACTCACATAGC
			gSPAex12m	CTTTACTAGACTAATTCTCCTG

13	1361 pb	4	gSPAex13a	CAGATTCAAGAAGACAGATC
			gSPAex13m	GCAATAATTCACCACACTTG
			gSPAex13n	GGTAGTTCTTGTTTCTGCTC
14	985 pb	4	gSPAex14a	CAAGTGTGGTGAATTATTGC
			gSPAex14m	GAGCTGAAAAGTATTCAGC
			gSPAex14n	TGCAAAGGACATAGCCAGTG
15	1076 pb	1	gSPAex15a	AGCCTCTGGAGATAGTATGC
			gSPAex15m	CTAGAACAGGGGTCACAGTC
			gSPAex15n	TTGGACTTCTTAAACTTC
16	1404 pb	4	gSPAex16a	GCAGTATGCAAGAAATTGAAC
			gSPAex16m	GGCCTGTAATTTTCTTCTG
			gSPAex16b	GTACTGAATAGATACATGTAG
17	445 pb	3	gSPAex17a	GTGTAGCAGATCAACATAG
			gSPAex17m	CATCTTCAAGTTTGGTGAC

Hormis l'exon 1, amplifié à l'aide du kit Advantage GC genomic PCR kit (Clontech) selon les instructions du fournisseur, quatre programmes de PCR légèrement différents (1, 2, 3 et 4) furent utilisés pour amplifier les exons de SPG4 (voir tableau 2). Les amplifications furent toutes effectuées dans un volume de 50 µl contenant 100 ng d'ADN génomique, 50 pmoles de chaque amorce, 250 µM de dNTPs, 1X de tampon Takara et 1 unité de Taq polymérase Takara La Taq (Shuzo Co.). Les réactions de PCR ont été réalisées selon le procédé du « hot start » : la Taq polymérase est ajoutée à 94°C après une première étape de dénaturation de 5 min à 96°C. Les échantillons sont par la suite soumis à 30 cycles de dénaturation (94°C pendant 40 sec), d'hybridation (prog. 1 : 60°C pendant 50 sec ; prog. 2, 58°C pendant 50 sec, prog. 3 et 4 : 55°C pendant 50 sec) et d'élongation (prog. 1 et 4 : 72°C pendant 1 min, prog. 2 et 3 : 72°C pendant 40 sec), suivis d'une dernière étape d'élongation (10 min à 72°C). Le séquençage de ces produits de PCR a été réalisé sur un séquenceur ABI 377 (PE-Applied Biosystems) en utilisant soit les amorces de PCR soit les amorces internes notées « b » et « n » (voir tableau 2).

4) Caractérisation de SPG4

Les clones d'ADNc 977312 (EST AA560327) et 568234 (EST AA107866) issus des banques d'ADNc de blastocyste et d'embryon E8 de souris, qui tous deux

correspondent à l'orthologue murin de SPG4, ont été isolés par le consortium IMAGE et séquencés au laboratoire sur un séquenceur ABI 377 (PE-Applied Biosystems). Afin d'analyser le profil d'expression de SPG4 et de son orthologue murin Spg4, les collections d'ADNc de différents tissus humains fœtaux et adultes, ainsi que de tissus de souris (panels MTC, Clontech) ont été testées par PCR selon le protocole du fournisseur avec le couple d'amorces SPA_Ca/SPA_Cm pour les ADNc humains et le couple SPA_Ca /spam (spam : 5'-ACCGAAGTCAAGAGCCTATC-3') pour les ADNc de souris. Les conditions de PCR sont celles utilisées pour l'amplification de SPG4 à partir d'ADNc de lignées de lymphoblastes (cf. § Détection des mutations), excepté que les échantillons ont été soumis à 32 cycles pour les ADNc issus de tissus adultes humains et de tissus murins, et à 28 cycles pour les ADNc issus de tissus fœtaux. Les produits d'amplification ont migré par électrophorèse sur des gels d'agarose 2 %.

5) Analyse histologique d'une biopsie musculaire d'un patient

Les analyses histologiques et histo-enzymatiques ont été réalisées à partir d'une biopsie musculaire d'un patient issu d'une famille liée au locus SPG4 selon les techniques standards décrites dans Casari et al. (17).

6) Numéros d'accès dans les bases de données publiques

L'ADNc SPG4 (ou SPAST) et la séquence protéique déduite, GenBank/EMBL AJ246001 ; le clone d'ADNc incomplet Spg4, GenBank/EMBL AJ246002 ; le gⁿ SPG4 (ou SPAST), GenBank/EMBL AJ246003.

Exemple 2 : Analyse de la séquence de l'intervalle SPG4

L'analyse des événements de recombinaison a permis de réduire la région candidate SPG4 à un intervalle génétique de 0 cM entre les marqueurs D2S352 et D2S2347 (19, 20). Une carte de préséquencage de l'intervalle SPG4 composé de 37 BACs a été construite (Hazan et al., sous presse dans Genomics) ; la région candidate couvre une distance physique d'environ 1,5 Mb. Douze BACs chevauchants, s'étendant sur l'intervalle SPG4 à l'exception d'un unique trou de 4 kb entre les clones A et E, ont été sélectionnés pour être séquencés (Fig. 1A). Sept de ces BACs (A, B, C, D, E, F et G), couvrant approximativement 70 % de la région d'intérêt, ont déjà été séquencés. Les séquences de ces 7 BACs ont été comparées à celles des bases de données nucléiques et protéiques, et analysées avec quatre programmes de prédiction d'exons. Ces analyses de séquences préliminaires ont permis de mettre en évidence 14 unités de transcription potentielles, dont trois correspondant aux gènes codant pour la xanthine deshydrogénase, la stéroïde 5 α -réductase 2 et une protéine liant le TGF β . Sur les 14 gènes détectés par l'analyse de séquence, 9 avaient été préalablement

identifiés dans les bases de données d'EST (pour « Expressed Sequence Tag ») et localisés au sein de l'intervalle SPG4 (Hazan et al., sous presse dans Genomics) ; les 5 gènes restants n'ont pu être identifiés qu'en séquençant la région candidate. L'un de ces 5 nouveaux gènes présentait une homologie en 3' de sa région codante avec les gènes codant pour la famille protéique des AAA (16). Des analyses de séquence plus approfondies ont montré que ce gène, nommé SPG4 (ou SPAST), était composé de 17 exons et s'étendait sur une région d'environ 90 kb, couverte par deux clones de BAC adjacents, D et G (cf. Fig. 1B). Les trois premiers exons prédits de ce gène furent identifiés dans le BAC D par deux des quatre programmes de prédiction d'exons utilisés, GRAIL II et GENSCAN ; ils présentent une forte homologie avec un EST de blastocyste de souris, AA560327. Les 14 derniers exons se trouvent dans le BAC G. La séquence protéique déduite des exons 7 à 17 est significativement homologue à une sous-classe de la famille des AAA, comportant les protéines de levure Yta6p (21), TBP6 (21) et End 13, ainsi que la protéine de souris SKD1 (22).

Sur les quatre programmes de prédiction d'exons, FGENEH semble le plus fiable et le plus puissant, permettant la détection de la plupart des gènes de cette région chromosomique en 2p21-p22. Cette constatation s'applique également au gène SPG4 pour lequel 15 exons ont pu être mis en évidence par ce programme quand seuls 4, 9 ou 11 exons ont pu respectivement être localisés par les programmes Genie, GRAIL II et GENSCAN. L'organisation génomique de ce gène (Fig. 1B) a pu être confirmée par la suite grâce à la détermination de la séquence de l'ADNc SPG4. Les jonctions introns/exons sont représentées sur le tableau 3 ci-après : la taille des exons varie de 41 pb (exon 16) à 1,410 kb (exon 17), celle des introns variant de 140 pb (intron 11) à 23,247 kb (intron 1).

Tableau 3
Organisation intron/exon du gène SPG4

Exon/ intron	Taille de l'exon (pb)	Position sur l'ADNc	Site accepteur d'épissage	Site donneur d'épissage	Taille de l'intron (pb)
1	540	1		TGAGAAAG/gtaactagggggctgg	23 247
2	87	541	atttttatttttaaag/CAGGACAG	AGGACAAAG/gtaagattgtattgt	1 943
3	84	628	aattttttctttcag/GTGAACAG	ACTTCTAG/gtatcaattaatgtat	9 190
4	96	712	cttcctgttgcatag/AGAAAGATG	CCAGTCAG/gtgggttaggtaac	15 745
5	188	808	acttttctctgtcag/AAAGTGGA	CTCATAAAG/gtattctgggacagta	876
6	134	996	tttgtatcctttaag/GGTACTCC	GTGGACAA/gtaagtttgccatct	283
7	94	1 130	aggctgtgttcttag/TGGAACAG	GGCCTGAG/gtaagaacttatatt	10 735
8	75	1 224	agtatatatttttag/TTGTTTAC	CAATGCTG/gtaagggtctctca	1 385
9	72	1 299	ctgtgatttttaaag/GCTAAAGC	CAAAATAC/gtgagtgctctgttc	8 083
10	76	1 371	taatgcttgttttag/GTGGGAGA	TTTTATAG/gtaagaacatatttc	238
11	92	1 447	ctgtatttcctctag/ATGAAAGTT	TTGATGGT/gtaagtggtgattatg	140
12	80	1 539	gatttttgctgtag/GTACAGTC	GTTCTCAG/gtagggagatttatat	4 715
13	43	1 619	ggatttttttttag/GCGTTTCA	ATGAGGAG/gtatgtatctgtgtt	1 389
14	80	1 662	tttaatatttttcag/ACAAAGACT	CTTGCTAG/gtgagtaattggatt	1 521
15	71	1 742	tccttcctctcctag/AATGACTG	TATCCGAG/gtaggtatacaagagc	2 210
16	41	1 813	ctttatgtttttacag/AACTAAAA	CCAGTGAG/gtatagtattttacaa	7 115
17	1 410	1 854	cttttaaaaatctag/ATGAGAAA		

Les séquences des exons et des introns sont respectivement indiquées en majuscules et minuscules.

Exemple 3 : Identification de l'ADNc SPG4

Plusieurs amplifications successives par RACE-PCR en 5' et 3' furent réalisées sur des collections d'ADNc de cerveau et de foie adultes et de cerveau fœtal, afin de caractériser le transcrit SPG4. Toutes les RACE-PCR en 5' ont donné des produits d'amplification se terminant à la position nt 263 de l'ADNc SPG4 (Fig. 2), ce qui était probablement dû au contenu riche en GC de la région 5' du transcrit (90 % de GC dans les 60 pb précédant la position nt 263). Quatre produits de PCR chevauchants, recouvrant la totalité de la région codante, ont été amplifiés à partir des ADNc issus des lymphoblastes de six individus contrôles et séquencés intégralement dans le but de vérifier la séquence du transcrit SPG4. L'alignement des séquences de tous les produits de PCR et RACE-PCR a permis de reconstituer une séquence de 3263 pb comprenant une phase ouverte de lecture de 1848 pb précédée par une région 5' non traduite (5' UTR pour « 5' UnTranslated Region ») de 125 pb et suivie par une région 3' UTR de 1290 pb incluant un site de polyadénylation entre les positions nt 3227-3232, ~ 35 pb en amont de la queue polyA (Fig. 2). La comparaison de la séquence de l'ADNc SPG4 avec les banques de données d'ESTs a permis de détecter une homologie significative avec 6 ESTs humains dont l'EST N47973 qui contient une région 3' non codante plus étendue (+ 180 pb) comprenant un deuxième site de polyadénylation. Le site d'initiation de la traduction a été identifié par la présence d'une séquence consensus de Kosak (CTGTGAatgA) définie comme un «contexte adéquat» à l'initiation de la traduction attendu qu'une purine est localisée 3 nt en amont de l'ATG initiateur, lui-même précédé d'un codon STOP. La séquence de l'ADNc de 3263 pb est identique à la séquence transcrite déduite des 17 exons du gène SPG4. L'analyse de la séquence de la région 5' à l'aide des programmes informatiques TSSG et TSSW suggère la présence d'une séquence promotrice de type TATA box située 43 pb en amont de la position nt 1 de l'exon 1.

Exemple 4 : Mutations dans le gène SPG4

Des mutations hétérozygotes ont été recherchées dans l'ADNc SPG4 provenant de lymphoblastes de 14 patients issus de familles liées au locus SPG4 (1 individu atteint par famille). Quatre fragments de PCR chevauchants I, II, III et IV recouvrant la phase ouverte de lecture de l'ADNc SPG4 ont été amplifiés et séquencés chez les 14 patients ainsi que chez 6 individus sains contrôles. L'électrophorèse sur gel agarose du fragment de PCR IV a montré trois bandes d'intensité égale chez 3 patients des familles 2992, 5226 et 5330 provenant de la même région de Suisse, ce qui suggérerait une microdélétion ou une mutation d'un site d'épissage ; les deux bandes

supplémentaires n'étaient pas présentes chez 2 individus sains issus des familles 2992 et 5330 (Fig. 3A). La séquence génomique de l'exon 16 a révélé une mutation hétérozygote A->G du site accepteur d'épissage (AG) de l'intron 15 chez les individus atteints de ces trois familles (Fig. 3B) ; cette mutation engendre la perte de l'exon 16 suivie d'un décalage de la phase de lecture dans le transcrit anormal. Aucun des membres sains incluant maris et épouses ne porte cette mutation du site d'épissage. L'identification de la même mutation chez tous les membres atteints de ces trois familles suisses démontre l'existence d'un ancêtre commun, ce qui avait été préalablement suggéré par l'étude des haplotypes.

- 10 Trois mutations ponctuelles 1210C->G, 1468G->A et 1620C->T qui introduisent des substitutions d'un acide aminé dans la séquence protéique (S362C, C448Y et R499C) ont été respectivement mises en évidence par le séquençage des fragments de PCR III et IV chez les individus atteints des familles 624, 4014 et 618. Ces trois substitutions impliquent toutes un résidu cystéine, induisant la perte ou l'insertion d'une
- 15 cystéine dans la séquence protéique. Une délétion de 1 pb, 1520delT, qui crée l'apparition d'un codon STOP induisant une protéine tronquée composée de 465 acides aminés (aa) a été détectée chez les individus atteints de la famille A. Aucune des cinq mutations résumées dans le tableau 4 ci-après n'a été trouvée chez les individus contrôles testés, qu'ils appartiennent à la fratrie saine ou aux conjoints des
- 20 sept familles analysées ici. Ces cinq mutations affectent de façon importante la séquence protéique dans un domaine très conservé, ou cassette AAA (23), qui est composé de plusieurs motifs protéiques supposés être responsables de l'activité ATPase chez tous les membres de la famille des AAA.

Tableau 4
Mutations dans SPG4 chez les patients atteints de PSF-AD

Famille	Localisation	Mutation ^a	Changement d'acide aminé ^b	Conséquence
624	exon 7	1 210 C → G	S362C	faux sens
4 014	exon 11	1 468 G → A	C448Y	faux sens
A	exon 11	1 520 delT	466STOPcodon	non sens
618	exon 13	1 620 C → T	R499C	faux sens
2 992	intron 15	1 813-2a → g	Δ aa564 → aa576 (PTC+7 aa)	perte de l'exon 16 + décalage
5 226	intron 15	1 813-2a → g	Δ aa564 → aa576 (PTC+7 aa)	perte de l'exon 16 + décalage
5 330	intron 15	1 813-2a → g	Δ aa564 → aa576 (PTC+7 aa)	perte de l'exon 16 + décalage

^a Les positions en nt font référence à la séquence de l'ADNc SPG4.

^b Les positions en aa font référence à la séquence de la spastin.

Les bases des exons sont indiquées en majuscules, celles des introns en minuscules.

PTC+7 aa = "premature termination codon" à 7 aa en aval de l'exon 16.

En addition à ces cinq mutations décrites précédemment, des recherches de mutations hétérozygotes réalisées sur des patients atteints de PSF-AD issus de 36 autres familles ont permis de mettre en évidence 34 autres mutations altérant ou susceptibles d'altérer le produit d'expression du gène SPG4.

- 5 Les caractéristiques de ces 34 autres mutations sont résumées dans le tableau 5 ci-après dans lequel ont été également insérées les cinq premières mutations précédemment citées.

Tableau 5
Mutations dans SPG4 chez les patients atteints de PSF-AD

Famille	Localisation	Mutation ^a	Changement d'acide aminé ^b	Conséquence
624	exon 7	1210 C→G	S362C	faux sens
6958	exon 8	1233 G→A	G370R	faux sens
214	exon 8	1267 T→G	F381C	faux sens
1002	exon 8	1283 T→G	N386K	faux sens
027	exon 8	1288 A→G	K388R	faux sens
019	exon 10	1401 C→G	L426V	faux sens
4014	exon 11	1468 G→A	C448Y	faux sens
148	exon 11	1504 G→T	R460L	faux sens
618	exon 13	1620 C→T	R499C	faux sens
636	exon 15	1788 G→A	D555N	faux sens
627	exon 15	1792 C→T	A556V	faux sens
2971	exon 3	702 C→T	Q193STOP	non sens
3655	exon 5	873 A→T	K229STOP	non sens
1010	exon 5	907 C→A	S261STOP	non sens
3938	exon 5	932 C→G	Y269STOP	non sens
6922	exon 10	1416 C→T	R431STOP	non sens
616	exon 10	1416 C→T	R431STOP	non sens
605	exon 15	1809 C→T	R562STOP	non sens
030	exon 2	578-579insA	PTC + 2 aa	décalage + non sens
615	exon 5	852del11	PTC + 18 aa	décalage + non sens
042	exon 5	882-883insA	PTC + 12 aa	décalage + non sens
032	exon 5	906delT	PTC + 17 aa	décalage + non sens
189	exon 9	1299delG	PTC + 3 aa	décalage + non sens
3686	exon 9	1340del5	PTC + 35 aa	décalage + non sens
625	exon 9	1340del5	PTC + 35 aa	décalage + non sens
A	exon 11	1520delT	PTC + 7 aa	décalage + non sens
115	exon 12	1574delGG	PTC + 2 aa	décalage + non sens
3266	exon 13	1634del22	PTC + 18 aa	décalage + non sens
149	exon 14	1684-1685insTT	PTC + 9 aa	décalage + non sens
645	exon 14	1685del4	PTC + 7 aa	décalage + non sens
029	intron 4	808-2 a→g	?	mutation site d'épissage
162	intron 6	1129+2 t→g	?	mutation site d'épissage
125	intron 7	1223+1 g→t	?	mutation site d'épissage
143	intron 8	1299+1 g→a	?	mutation site d'épissage
1620	intron 11	1538+5 g→a	(PTC + 6 aa)	perte de l'exon 11 + décalage
1006	intron 11	1538+3 del4	?	mutation site d'épissage
1605	intron 13	1661+1 g→t	?	mutation site d'épissage
1012	intron 13	1662-2 a→t	?	mutation site d'épissage
1626	intron 15	1812+1 g→a	?	mutation site d'épissage
2992	intron 15	1813-2 a→g	Δ aa564 → aa576 (PTC+7 aa)	perte de l'exon 16 + décalage
5226	intron 15	1813-2 a→g	Δ aa564 → aa576 (PTC+7 aa)	perte de l'exon 16 + décalage
5330	intron 15	1813-2 a→g	Δ aa564 → aa576 (PTC+7 aa)	perte de l'exon 16 + décalage
1611	intron 16	1853+1 g→a	?	mutation site d'épissage

^a Les positions en nt font référence à la séquence de l'ADNc SPG4. ^b Les positions en aa font référence à la séquence de la spastin. Les bases des exons sont indiquées en majuscules, celles des introns en minuscules. PTC+n aa - "premature termination codon" à n acide aminé en aval de la mutation.

Exemple 5 : Analyse de la séquence protéique de la spastin

La phase ouverte de lecture de SPG4 code pour une protéine de 616 aa que nous avons nommée spastin et dont le poids moléculaire est d'environ 67,2 kDaltons (kD). La comparaison de cette séquence en acides aminés dans les bases de données protéiques à l'aide des programmes BLAST a permis de mettre en évidence une zone de forte homologie avec plusieurs membres de la famille des AAA à l'extrémité C-terminale de la spastin. Les motifs « types » de la famille des AAA, englobés dans la cassette AAA, sont localisés entre les positions aa342 et aa599 (voir Fig. 2) d'après les comparaisons de séquence dans les bases de données de domaines protéiques ProDom et Prosite. Les trois domaines types conservés, dont les motifs A et B de Walker ainsi que le motif consensus minimal des protéines AAA sont respectivement situés au sein de la cassette AAA aux positions aa382-389, aa437-442 et aa480-498 (Fig. 2). Le motif A de Walker, « GPPGNGKT » appelé également p-loop (ou boucle-p) qui correspond au domaine de fixation de l'ATP et le motif B « IIFIDE » sont très conservés entre tous les membres de la famille des AAA incluant la spastin.

La comparaison des cassettes AAA présentes dans 150 protéines de cette famille d'ATPase issues d'organismes très éloignés dans l'évolution a permis de classer cet ensemble de protéines en plusieurs sous-groupes, en fonction du nombre de cassettes AAA identifiés (1 ou 2) et des homologies de séquence entre ces différentes cassettes (23). Parmi toutes les protéines de la famille des AAA, la spastin présente une plus forte homologie avec une sous-classe particulière des AAA, et plus spécifiquement avec les protéines suivantes dont la plupart ont été identifiées grâce au séquençage complet du génome de l'organisme considéré : deux protéines de *Caenorhabditis elegans* O16299 et Q18128, deux sous-unités du protéasome 26S de *Saccharomyces cerevisiae* Yta6p (Q02845) et TBP6 (P40328) (21), une sous-unité du protéasome de *Schizosaccharomyces pombe* (O43078), les protéines SAP1 (P39955) et END13 (P52917) de *S. cerevisiae* et la protéine murine SKD1 (P46467) (22). L'alignement multiple de ces 8 protéines avec la spastin est représenté sur la Fig. 4A. Sur les 257 acides aminés englobant la cassette AAA (positions aa342-599), la spastin présente une identité de séquence de 52 %, 51 % et 50 % avec la protéine de levure Yta6p (Q02845), la protéine de nématode O16299 et la protéine de levure TBP6 (P40328) respectivement. Des résultats similaires ont été obtenus par l'analyse de la séquence protéique de la spastin dans la base de données ProDom qui a montré l'existence de trois domaines d'homologie (nommés 92, 179 et 6226 et correspondant aux positions aa342-409, aa411-509 et aa512-599) trouvés dans les sous-unités

putatives du protéasome 26S de levure. En outre, les membres de ce sous-groupe des AAA contiennent le plus souvent des motifs de type leucine-zipper dont deux ont pu être détectés dans la séquence protéique de la spastin aux positions aa50-78 et aa508-529 par l'analyse de la séquence dans la base de données Prosite (voir Fig. 2).

- 5 Cette analyse a également pu prédire la présence d'un motif de dimérisation de type hélice-boucle-hélice (« helix-loop-helix ») situé entre les positions aa478 et aa486.

- La comparaison de la séquence protéique de la spastin avec celles des métalloprotéases mitochondriales comme les protéines de levure AFG3, RCA1 et YME1, ainsi que la paraplegin qui est impliquée dans une forme rare de PSF-AR
- 10 montre que l'homologie entre ces cinq membres de la famille des AAA est limitée à la région de 257aa englobant la cassette AAA (Fig. 4B). Dans cette région, l'identité de séquence entre la spastin et la paraplegin n'est que de 29 % alors que la paraplegin et la protéine de levure AFG3 sont identiques à 57 % sur cette même portion de la séquence protéique. Cette comparaison de séquence suggère que la spastin
- 15 n'appartient pas au même sous-groupe des AAA que la paraplegin et autres métalloprotéases mitochondriales. De plus, l'analyse informatique de la séquence de la spastin avec le programme PSORT II qui permet de prédire la localisation sub-cellulaire des protéines semble indiquer que le spastin est une protéine nucléaire. Un éventuel signal de localisation nucléaire (NLS pour « Nuclear Localization Signal »),
- 20 RGKKK, a été mis en évidence entre les positions aa7 et aa11 alors qu'aucun peptide signal caractéristique d'un import dans la mitochondrie n'a pu être décelé, contrairement à ce qui avait été observé pour la paraplegin.

Exemple 6 : Profils d'expression de SPG4 et de son orthologue murin Spg4

- La comparaison de la séquence nucléique de SPG4 dans les banques de
- 25 données d'EST a permis de détecter plusieurs ESTs humains, murins et de rat présentant une forte homologie avec SPG4. Les clones d'ADNc de blastocyste et d'embryon E8 de souris correspondant à deux des ESTs murins, AA560327 et AA107866, ont été obtenus du consortium IMAGE et séquencés intégralement. L'assemblage des séquences de ces clones d'ADNc a permis de reconstituer une
- 30 séquence consensus de 1689 pb incluant une phase ouverte de lecture incomplète de 1514 pb. La comparaison entre l'ADNc SPG4 humain et cet ADNc de souris a montré qu'il manque au transcrit murin environ 460 pb à l'extrémité 5' dont le codon d'initiation de la traduction. La phase de lecture ouverte de souris est suivie d'une région 3' non codante (3' UTR) de 175 pb contenant un site de polyadénylation situé à ~20 pb en
- 35 amont de la queue polyA (Fig. 5). La séquence nucléique de SPG4 et la séquence

protéique de la spastin humaine présentent respectivement une identité de 89 % (entre les positions nt 460 et nt 1982) et de 96 % (entre les positions aa113 et aa616) avec les séquences de l'ADNc et de la protéine déduite de la souris. Ce degré important d'homologie permet d'affirmer que ce transcrit de souris correspond à l'orthologue murin de SPG4, qui a été donc baptisé Spg4.

L'hybridation de northern blots comprenant les ARNm de divers tissus humains et murins (Clontech) avec les clones d'ADNc SPG4 et Spg4 n'a pas donné de résultats probants excepté une très faible bande correspondant à un transcrit de 2,5 kb dans le testicule de souris après 10 jours d'exposition. En raison du faible niveau d'expression de ce gène, les profils d'expression de SPG4 et Spg4 ont été déterminés par des expériences de PCR sur des collections d'ADNc normalisées provenant de divers tissus adultes et fœtaux (voir Fig. 6). Le gène murin Spg4 est exprimé de façon ubiquitaire dans les tissus adultes de souris ainsi que du stade E7 au stade E17 de l'embryon de souris (Fig. 6A). Une plus forte expression de Spg4 a été détectée dans le foie, le muscle squelettique et les testicules, ainsi qu'au stade E15 de l'embryon. L'expression précoce de Spg4 au cours du développement embryonnaire a été confirmée par la présence d'ESTs provenant de banques d'ADNc de blastocyste, d'embryon E8 et de carcinome embryonnaire dans les banques de données publiques d'ESTs. Le gène humain SPG4 est lui aussi exprimé de façon ubiquitaire dans les tissus adultes (Fig. 6B) et fœtaux (Fig. 6C), avec une expression peut-être plus marquée dans le cerveau fœtal.

Exemple 7 : Pas de défaut de la phosphorylation oxydative dans la PSF-AD liée au locus SPG4

Afin de déterminer si des mutations de la spastin induisaient un défaut de la phosphorylation oxydative (OXPHOS) dans la mitochondrie, à l'image de ce qui avait été observé pour la paraplegin, une biopsie musculaire a été réalisée sur un patient d'une des familles de PSF-AD liée au locus SPG4. Les analyses morphologiques et histo-enzymatiques de cette biopsie de muscle n'ont pas révélé de fibre musculaire de type RRF (pour « ragged red fibers »), caractéristique des défauts OXPHOS dans la mitochondrie. Le fait que toutes les fibres musculaires apparaissent normales ainsi que la prédiction d'une localisation nucléaire de la spastin semblent indiquer que la PSF-AD liée au locus SPG4 n'est pas une maladie mitochondriale de type OXPHOS, par opposition à la PSF-AR liée au locus SPG7.

Par une approche de clonage positionnel basée sur le séquençage d'une région de 1,5 Mb, nous avons identifié le gène SPG4 (ou SPAST) responsable de la forme la plus fréquente de PSF-AD, préalablement localisé sur les bandes chromosomiques 2p21-p22. Trente-neuf mutations altérant ou susceptibles d'altérer le produit du gène, nommé spastin, ont pu être détectées chez les individus atteints de quarante-et-une familles de PSF-AD présentant une liaison au locus SPG4. La spastin est un nouveau membre de la famille des protéines AAA, dont la localisation semble être nucléaire et qui présente une forte homologie avec les sous-unités du protéasome 26S de levure. En dépit d'une grande homologie restreinte à un domaine de 230 à 250 aa, dite cassette AAA, les nombreux membres de cette famille protéique peuvent participer à des mécanismes cellulaires très variés comme le transport de protéines au sein de vésicules, la régulation du cycle cellulaire, la biogénèse des organelles, le contrôle de la transcription, Toutefois, tous ces mécanismes cellulaires impliquent l'assemblage, la fonction ou la dégradation de complexes protéiques, ce qui suggère que les membres de la famille des AAA sont des protéines dites « chaperons ».

Références

1. Reid, E. Pure hereditary spastic paraplegia. *J. Med. Genet.* 34, 499-503 (1997).
2. Bruyn, R.P.M. & Scheltens, P.H. Hereditary spastic paraparesis (Strumpell-Lorain) in Handbook of clinical neurology Vol. 15 (ed. de Jong, J.M.B.V.) 301-318 (Elsevier Science Publishers B.V., 1991).
3. Werderlin, L. Hereditary ataxias. Occurrence and clinical features. *Acta Neurol. Scand.* 73 (Suppl. 106) (1986).
4. Polo, J.M., Calleja, J., Combarros, O. & Berciano, J. Hereditary ataxias and paraplegias in Cantabria, Spain. An epidemiological and clinical study. *Brain* 114, 855-866 (1991).
5. Skre, H. Hereditary spastic paraplegia in Western Norway. *Clin. Genet.* 6, 165-183 (1974).
6. Hazan, J., Lamy, C., Melki, J., Munnich, A., de Recondo, J., & Weissenbach, J. Autosomal dominant familial spastic paraplegia is genetically heterogeneous and one locus maps to chromosome 14q. *Nature Genet.* 5, 163-167 (1993).
7. Hazan, J. et al. Linkage of a new locus for autosomal dominant familial spastic paraplegia to chromosome 2p. *Hum. Mol. Genet.* 3, 1569-1573 (1994).
8. Hentati, A. et al. Linkage of a locus for autosomal dominant familial spastic paraplegia to chromosome 2p markers. *Hum. Mol. Genet.* 3, 1867-1871 (1994).
9. Fink, J.K. et al. Autosomal dominant familial spastic paraplegia : tight linkage to chromosome 15q. *Am. J. Hum. Genet.* 56, 188-192 (1995).
10. Hedera, P. et al. Novel locus for autosomal dominant hereditary spastic paraplegia, on chromosome 8q. *Am. J. Hum. Genet.* 64, 563-569 (1999).
11. The Hereditary Spastic Paraplegia Working Group. Hereditary spastic paraplegia : advances in genetic research. *Neurology* 46, 1507-1514 (1996).
12. Durr, A. et al. Phenotype of autosomal spastic paraplegia linked to chromosome 2. *Brain* 119, 1487-1496 (1996).
13. Nielsen, J.E. et al. CAG repeat expansion in autosomal dominant pure spastic paraplegia linked to chromosome 2p21-24. *Hum. Mol. Genet.* 6, 1811-1816 (1997).
14. Jouet, M. et al. X-linked spastic paraplegia (SPG1), MASA syndrome and X-linked hydrocephalus result from mutations in the L1 gene. *Nature Genet.* 7, 402-407 (1994).
15. Saugier-Verber, P. et al. X-linked spastic paraplegia and Pelizaeus-Merzbacher disease are allelic disorders at the proteolipid protein locus. *Nature Genet.* 6, 257-262 (1994).

16. Confalonieri, F. & Duguet, M. A 200-amino acid ATPase module in search of a basic function. *BioEssays* 17, 639-650 (1995).
17. Casari, G. et al. Spastic paraplegia and OXPHOS impairment caused by mutations in Paraplegin, a nuclear-encoded mitochondrial metalloprotease. *Cell* 93, 973-983 (1998).
18. Osoegawa, K. et al. An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 52, 1-8 (1998).
19. Scott, W.K. et al. Locus heterogeneity, anticipation, and reduction of the chromosome 2p minimal candidate region in autosomal dominant familial spastic paraplegia. *Neurogenetics* 1, 95-102 (1997).
20. Heinzlef, O. et al. Mapping of a complicated familial spastic paraplegia to locus SPG4 on chromosome 2p. *J. Med. Genet.* 35, 89-93 (1998).
21. Schnall, R. et al. Identification of a set of yeast genes coding for a novel family of putative ATPases with high similarity to constituents of the 26S protease complex. *Yeast* 10, 1141-1155 (1994).
22. Perier, F. et al. Identification of a novel mammalian member of the NSF/CDC48p/Pas1p/TBP-1 family through heterologous expression in yeast. *FEBS lett.* 351, 286-290 (1994).
23. Beyer, A. Sequence analysis of the AAA protein family. *Protein Sci.* 6, 2043-2058 (1997).
- Barany, F., (1991), *Proc. Natl. Acad. Sci. USA*, 88, 189-193.
- Bodansky M., *Principles of peptide synthesis*, (1984).
- Buckholz, R.G. *Curr. Op. Biotechnology* 4 : 538-542, 1993.
- Burg, J.L. et al. (1996), *Mol. and Cell. Probes*, 10, 257-271.
- 25 Carter, B.J. *Curr. Op. Biotechnology* 3 : 533-539, 1993.
- Cherif D., Julier, C., Delattre, O., Derré, J., Lathrop, G.M., and Berger, R. *Proc. Natl. Acad. Sci. USA*. 87 : 6639-6643, 1990.
- Chu, B.C.F. et al. (1986), *Nucleic Acids Res.*, 14, 5591-5603.
- Chumakov, I., Rigault, P., Guillou, S., Ougen, P., Billault, A., Guasconi, G., Gervy, P., Le Chumakov, I.M., Rigault, P., Le Gall, I., et al. *Nature* 377 : 175-183, 1995.
- 30 Duck, P. et al. (1990), *Biotechniques*, 9, 142-147.
- Edwards, C.P., and Aruffo, A. *Curr. Op. Biotechnology* 4 : 558-563, 1993.
- Epstein, A. *Médecine/Sciences* 8 : 902-911, 1992.
- Erich, H.A., (1989), New York : Stockton Press.

- Guatelli J.C. et al. *Proc. Natl. Acad. Sci. USA* 87: 1874-1878, 1990 et al. *Cell* 85 : 281-290, 1996.
- Innis, M.A. et al. (1990), Academic Press.
- Kievitis, T. et al. (1991), *J. Virol. Methods*, 35, 273-286.
- 5 Köhler et Milstein. *Nature* 256, 495-497, 1975.
- Kwoh, D.Y. et al. (1989), *Proc. Natl. Acad. Sci. USA*, 86, 1173-1177.
- Landegren U., Kaiser R., Sanders J. & Hood L. *Science* 241 : 1077-1080, 1988.
- Lizardi, P.M. et al. (1988), *Bio/technology*, 6, 1197-1202.
- Luckow, V.A. (1993), *Curr. Op. Biotechnology* 4, 564-572.
- 10 Matthews, J.A. et al. (1988), *Anal. Biochem.*, 169 : 1-25.
- Miele, E.A. et al. (1983), *J. Mol. Biol.*, 171 : 281-295.
- Olins, P.O., and Lee, S.C. *Curr. Op. Biotechnology* 4 : 520-525, 1993.
- Perricaudet, M., Stratford-Perricaudet, L. and Briand, P. *La Recherche* 23 : 471-473, 1992.
- 15 Rohlmann, A., Gotthardt, M., Willnow, T.E., Hammer, R.E., and Herz, J. *Nature Biotech.* 14 : 1562-1565, 1996.
- Rolfs, A. et al. (1991), Berlin : Springer-Verlag.
- Sambrook, J., Fritsch, E.F., and Maniatis, T. *Molecular cloning : a laboratory manual*. Sec. Ed. Cold Spring Harbor Lab., Cold Spring Harbor, New York.
- 20 Segev, D., (1992), Kessler C. Springer Verlag, Berlin, New-York, 197-205.
- Stone, B.B. et al. (1996). *Mol. and Cell. Probes*, 10 : 359-370.
- Stewart J.M. et Yound J.D., *solid phase peptides synthesis*, Pierce Chem. Company, Rockford, 111, 2ème éd., (1984).
- Suggs S.V., Wallace R.B., Hirose T., Kawashima E.H. and Itakura K. *PNAS* 78 : 6613-6617, 1981.
- 25 Temin, H.M. *Retrovirus vectors for gene transfer*. In Kucherlapati R., ed. *Gene Transfer*, New York, Plenum Press, 149-187, 1986.
- Walker G.T., Fraiser M.S., Schram J.L., Little M.C., Nadeau J.G., & Malinowski D.P. *Nucleic Acids Res.* 20 : 1691-1696, 1992.
- 30 Woo S.L.C. *Methods Enzymol.* 68 : 389, 1979.

Recherche scientifique

LISTE DE SEQUENCES

<110> CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE - CNRS

<120> CLONAGE, EXPRESSION ET CARACTERISATION DU GENE SPG4
RESPONSABLE DE LA FORME LA PLUS FREQUENTE DE PARAPLEGIE
SPASTIQUE AUTOSOMIQUE DOMINANTE

<130> D18374

<160> 71

<170> PatentIn Vers. 2.0

<210> 1

<211> 110000

<212> DNA

<213> Homo sapiens

<220>

<221> exon

<222> (9932)...(10471)

<220>

<221> intron

<222> (10472)...(33718)

<220>

<221> exon

<222> (33719)...(33805)

<220>

<221> intron

<222> (33806)...(35748)

<220>

<221> exon

<222> (35749)...(35832)

<220>

<221> intron

<222> (35833)...(45022)

<220>

<221> exon

<222> (45023)...(45118)

<220>

<221> intron

<222> (45119)...(60863)

<220>

<221> exon

<222> (60864)...(61051)

<220>

<221> intron

<222> (61052)...(61927)

<220>
<221> exon
<222> (61928)...(62061)

<220>
<221> intron
<222> (62062)...(62344)

<220>
<221> exon
<222> (62345)...(62438)

<220>
<221> intron
<222> (62439)...(73173)

<220>
<221> exon
<222> (73174)...(73248)

<220>
<221> intron
<222> (73249)...(74633)

<220>
<221> exon
<222> (74634)...(74705)

<220>
<221> intron
<222> (74706)...(82788)

<220>
<221> exon
<222> (82789)...(82864)

<220>
<221> intron
<222> (82865)...(83102)

<220>
<221> exon
<222> (83103)...(83194)

<220>
<221> intron
<222> (83195)...(83334)

<220>
<221> exon
<222> (83335)...(83414)

<220>
<221> intron
<222> (83415)...(88129)

<220>
<221> exon
<222> (88130)...(88172)

<220>
 <221> intron
 <222> (88173)...(89561)

<220>
 <221> exon
 <222> (89562)...(89641)

<220>
 <221> intron
 <222> (89642)...(91162)

<220>
 <221> exon
 <222> (91163)...(91233)

<220>
 <221> intron
 <222> (91234)...(93443)

<220>
 <221> exon
 <222> (93444)...(93484)

<220>
 <221> intron
 <222> (93485)...(100599)

<220>
 <221> exon
 <222> (100600)...(102009)

<400>	1							
taactgactc	tgctgtgttt	ccttggcatt	atagctaadc	aaattgagca	ggtcaggtaa			60
cagttttatac	ttacacctac	tattttcaaaa	ccatgagctc	attcacattt	tcactgaagt			120
aacaaatcct	ccataaacta	gaaaatctca	aactgggtgac	tgggagtttt	ggttttgttt			180
ttttgttggt	ttattttatt	ttattttatt	ttctagatgg	agtcttgctc	tggtgcccag			240
gctggaatgc	aatggcatga	tctcagctca	ctgcaacctc	cacctttcgg	gttcaagcaa			300
tcctcctgcc	tcaatcttcc	aagtagctgg	gactacagga	atgagctgcc	gcacctggcc			360
tggtttgttt	gtttttaaat	tttgaggcca	ggtgcagtgg	cccatatctg	tgatcccagc			420
actttgggag	accaaggcag	gccgattact	tgaggtcagg	agttcaagac	cagccaggcc			480
aacatggtaa	aaccatgtcg	ctactaaaaa	cacaaaaatt	aggccgggca	cggtggctca			540
cgtctataat	cccagcactt	tgggaggcca	aggtgagtgg	atcacctgag	gtcaggaggt			600
caagaccagc	ctggcaaaca	tggtgaaacc	ccgtctctat	taaaaataca	aaaattagcc			660
gggtgtggtg	gcgggcacct	gtaatcccag	ctattgggga	ggctgaggca	ggagaatcgc			720
ttgaacctgg	gaggtggagg	ttgcagtgg	tcgagatgg	gccactgacc	tccagcctgg			780

gcaataagag	cgagacgcca	tctcaaaaat	atataaataa	ataaataaaa	atacaaaaag	840
tagctaggct	ttgtggcagg	cacctgtagt	cccagctact	taggaggctg	aggcaggaga	900
attgcttgaa	cccaggaggc	ggaggttgca	gtgagccaag	attgtgccac	tgcgttccag	960
cctgggtgac	agagcaagac	tccttctcaa	aaataaataa	ataaataaat	aaataaataa	1020
ataaataaaa	ttttgagctg	ggcatgaaag	ctgaggcagg	aggatccctt	gagcccagca	1080
gtttgagacc	ccagtgaagt	ataattctga	cactgcactt	cagcctggct	gacagaggga	1140
gaccgtgtat	ctaaaaagaa	taaaataaca	atgatttttg	agccaataac	tcttagccaa	1200
tagatttcac	ataaaaattt	agaattctgg	tttctcttga	aaaattaaaa	aaatctgaca	1260
atgttaagct	tcacattcct	gaaaggcaaa	aatcagtgga	agctagctgg	gtgctgtggc	1320
tcacgcctat	aaacccagtg	ctttgggagg	ctagggtaag	tggattgggt	gagtccagga	1380
gtttgagatc	agcctggcca	acacagtcaa	accccatctc	tacaaaacat	acaaaaatta	1440
gctgggatat	ggtggtgtgc	acttgtggtc	ccaactactc	aggaggctga	ggcaggagga	1500
ccacttgagc	cctggagggt	gagggggcag	tgagccatga	ttgtgccact	gcactccagc	1560
ctggacgaca	gagagacatt	gtctcaaaaa	gaaaaaaaaa	aatcagctca	gtgggagctg	1620
agtaacagct	gtcctgttca	attacaggat	gcaactcttt	agcttctcat	agtttccatc	1680
attcacatca	tacatgcatt	tttgcatgcc	acacaacca	cacatggaac	ccatatatgt	1740
tatatgcctg	accactgttg	ctattggaag	ttttggccac	tgcattaaac	tataaactcc	1800
atcttattaa	tcctgacacc	ccacttattg	cctgatatat	tgtccgtttc	ttaatatcta	1860
ttcatagaac	aatgaatga	ataatatgtg	ccacattgtg	gactcaattc	agggagatga	1920
ctaattcatc	acaattatgt	ttttccttct	taatacagag	attggataat	aattccccaa	1980
ttgttacttc	tctcatccct	cccctccaac	catatctttt	catttgtttt	tcttatttta	2040
ggttttctgc	ctttttcaaa	atcagccatt	tcctcactgg	actctacatg	tgccattttt	2100
tttttttttt	ttttactaat	tttttttagt	gaaaagaggt	ccttaatatc	tgtcattggt	2160
ccacacttga	aatctagaat	ctctcgaatt	gaaggctctga	agagttcctt	agaaggcaag	2220
tagggccgaa	atatccaaaa	aaaactattt	gatggtaggc	actgtggtaa	ataaatatgg	2280
gttaaagtga	taaagtgacc	gaatgtcctg	gattagttag	cacagtacct	agctccttct	2340
ttttgtcctt	gtgaaactgt	ctcagattcc	attcaagatt	aagtgtcctg	aaagttctta	2400
caagctgaaa	actgggaggc	tctcacactg	taggtagaat	ggctagcagg	gttgggatct	2460
catctaggca	cattgcagaa	ccagggtaac	accaaggcat	tatttttttt	ctttctttac	2520
ttttcttctt	tttttttaga	gacagggtct	cattatgtta	cccaggctgg	attccaattc	2580

ctgagctcaa	gtgactcctcc	tgagtagctg	ggacttcagg	tgtgcaccgc	tgtgtctgaa	2640
ggcactgttt	tttttgtttt	ttttgttttt	tttgacacaa	atttaggaag	atgttaattc	2700
acaacagtct	tgagactgag	atataattcc	aaggagcaga	agatgtgagt	ttagagcact	2760
aaaaaatgag	attccagtag	agtcagaaat	ctgaaatggc	attacagata	taagagacaa	2820
aaacaaaatg	tattgagctc	tgtcatgttg	caggcatcat	gatggagggt	ttagatgtac	2880
tctttcattt	tgtaattttt	atagaggaat	taactagaat	agcaacccca	gtcccactaa	2940
ctctaaatcc	catattttta	ccatacaaaa	aaagagcaaa	agtgcagaaa	agcacagtca	3000
atattaagta	caaccagata	gcagagactc	agtaaattggg	aggccggagg	cccgaacccc	3060
aacataatgc	ccatgaatga	aagcccatca	cttgcgcttc	aggggctaac	aataacttta	3120
tttcataatt	aaaatagaaa	cagactatgt	aaaaatatta	ttcttgagat	cccagatttt	3180
taccttaa	tactgaggca	gtaagtgtaa	ttaactaata	tgtgatgttg	ggcaaataac	3240
agactcttag	agccccaaat	tctttatttt	aaaaaactga	ggccagatga	ggtggcttac	3300
acatgtaatc	ccagcactgg	ggatgccaaa	gcaggcagat	cacttgagct	tgagcttagg	3360
agtttgagat	cagcctgggc	aacacggtta	agacctcatc	tctaaacaaa	atacaaaaat	3420
tagccaaaca	tggtgggtgtg	cacctgtgtt	cctggctact	caggaggctg	aggtggaagg	3480
atcgcttgag	cccagagagg	tgagggtgca	gtgagccatg	atcttgccat	tgactcttac	3540
actctagtct	gggtgacaga	gtgagactct	gtcttggggg	aaaacaaaag	agatgataat	3600
gcttaactga	agtagcaata	ttttaaaaag	gcactaaaag	ttcatctgct	tagttcagaa	3660
tatgggttcg	ataaatatta	gcaagtagta	gtagtcatca	tcatcactgt	cactgctgtt	3720
ctctccttaa	acttaagcat	gttttttggt	tttttgagac	agtgtctcac	tctgtcacc	3780
aagctggagt	tcagtgggtgt	gatcttggct	cactccaacc	tctgcctccc	aggttcaagt	3840
gattctcctg	tctcaccctc	ccaagtagct	gggaccacag	acacgtacca	caaccacacc	3900
cagctaattt	ttcgtatttt	tggtagagg	ggggtttcac	catgttggcc	aggctggtct	3960
tgaactcctg	acctcaagtg	atccacctgt	ctcggcctcc	caaagtgtgt	ggattacagg	4020
cgtagccac	agcatccagc	cttaagcatg	ttaattaagt	ttttataatt	cagcaaaaatg	4080
gttggaat	gctgtcttaa	atgagatgct	taagctgccg	tctgaacatg	aggtagaagg	4140
aaattctaca	cataatcatt	gtgctaaaat	acttgcaaag	atggccacaa	caattcctcc	4200
tatcctcata	tatatgcccc	tttgcaatgt	gactttgcta	cttctctatc	aagatgtgga	4260
gcttattttc	ccatatattg	cactagagtt	ggccttctga	cttgctttga	caatggaatg	4320
tagtacaaat	gacactgtgc	aactttggat	tttaggtttc	gagagaactt	acaccttcca	4380

ctcacactct	cttggaacc	agatgcaatg	taaagaagtc	agggctatcc	tgctagagac	4440
atatgtccca	gctaatagcc	acaatcaacc	tctgaacata	tgaatgaggc	tagctaggcc	4500
atacagccat	tcggtcaagc	catcagatga	ctacatccac	aggaatgatc	cacaggcaag	4560
gccatcagaa	gaaccatcca	gctgaactta	ccccaaattg	ctgagtcaca	aagttgtgtg	4620
taaataaatg	tctgctatct	taagccagtg	agttttggag	tggtatatta	catagcatca	4680
gaaatctaac	acaatcatta	tgtttgaatc	atttttcaaa	tttctcatat	ttattaaatg	4740
agtaccataa	gcaaggtgtc	aggctggatg	caaaaagtga	ggcaaaatgt	ataaagtgtg	4800
accactgcct	tcagtaagtt	tacaatctat	atcaagaggt	gatgaagtgt	ttaaataatc	4860
atcctgcagg	gcaatatagt	ataagagcca	cagagtaaca	caaccatatt	gtcataacaa	4920
ctgaaaaaca	agatcatttc	tgctggaggt	gataatggaa	taatttatca	agaatataac	4980
agagctggac	gcggtggctc	acacctgtaa	tcccagcact	ttgggaggcc	aaggaaggtg	5040
gatcacaagg	tcaggagttc	gagaccatcc	tggtatacac	gatgaacccg	tctctactaa	5100
aaatacaaga	aattagccgg	gcgtggtggc	acgcgtctgt	agtcgcagct	actcaggagg	5160
ctgaggcagg	agaaccactt	gaacatggga	agcagagggt	gcattgagct	gagatcgtgc	5220
catggcactc	cagcctgggt	gacagagtga	gactcagtct	caaaaaaaaa	aaaaaaaaaa	5280
aaaatataac	attagaggta	agtcttgaag	gactttgaca	gtggaagtag	gaggcgaggc	5340
cattctaagt	gaatgaaaaa	tgacaggaga	gtaattgtag	tcctggaaaa	gagcaaagta	5400
ggtacagacc	aacagtctat	attagctaga	gtatagtga	agtgcagagg	aaatgtcgga	5460
gaaccattct	ttattcaaaa	actatcttcc	tcatggccaa	gcatagtggc	tcatgcctgt	5520
aatcccagca	ttttgggagg	tcaaggtgag	tggtactact	gagctcagga	attcaagacc	5580
atctggggca	acatagttag	acctcatctc	aactaaaaaa	caaaaaattc	agacagatgc	5640
agtggctcac	acctgtaatc	ccagaacttt	ggtaggctga	ggcgggcgga	tcacgaggtc	5700
aggagatcaa	gaccctcctg	gacaacatgg	agaaacccca	tctctattaa	aaatacaaaa	5760
ttagctgggc	atgggtggac	atccctgtaa	tcccagctac	tcgggaggct	gaggcaggag	5820
aatcgcttga	accagggagt	cggaggttgc	agtgagccga	gatcgcacca	ctgcactcca	5880
gtctggcgac	agagcgagac	tccatcttaa	aaaataaata	aattttaaaa	aaaactaccc	5940
cagcatggtg	gtgcatgcct	gtagtccag	ttactcagga	ggctgaggca	agagggtggt	6000
ttgagccagg	gaggtcaagg	ctgcagttag	ctctgatggc	gccactgtac	tccagcttgg	6060
gtgacagagt	gagaccttgt	ctcaaaaaca	aaaacaaaaa	caaaaaacca	acaaatctcc	6120
ttgttagtat	catggtgagt	aaaaaataaa	ataaaaatag	aaataaactg	aacatggtgg	6180

ctcatgcctg	taatcctagc	actttggaag	gctgaagtgg	gaggattgct	tgagggctgg	6240
agttcaaaac	tggcttgggc	aacacggtga	gagagacctt	gtctctacaa	aagaactttt	6300
aaaacaaaaa	atagataatt	taaaaaaatt	aaaaaaaaca	aaaaataaaa	aaataatcaa	6360
gtatcaactt	gattccaggc	actgcttact	actctagtgt	tatactgtag	atgtggaagc	6420
tgagtaactc	atccaagatc	accgaaagtg	atggaacaca	gatctaaatg	caaccagtct	6480
gactccagga	ccatttaacc	attctactat	tgggccctat	cttggctaag	ttagaaagta	6540
agttactttc	tttagtggtg	aagactggag	ggataacagg	gaagatagtt	atttaagaaa	6600
aaaaactggc	atcaaaactaa	atatccatca	atagttgaac	agtaaaatag	gttgtggtaa	6660
attcatataa	tgggaatacta	tatagcagtg	aaaatgtacc	acagttatag	aatcaacag	6720
ggaggaatth	caacacttaa	ttattaagta	ggtagccagg	catagcgggt	tatgcctgta	6780
atcccagcac	tttgggagac	caagacagga	ggattacttg	agcccagggg	ttcgagatca	6840
acctgggcaa	cagtgagact	ccatctctat	tttcttaaaa	taaaataaat	gaaatthtaa	6900
aaatthtgag	gagggaaaagc	aaacaaggga	tacttgaaat	atgattacat	ttccataaag	6960
tcaaagtgag	gcaaaatcat	acaagacatt	gtttagaaat	acataaatac	actgcaaact	7020
aaaaatgaga	cactagaatg	attaatataa	aattcaggat	agtggcttcc	tctagaggaa	7080
gagacaagac	attgagatta	gggaggagct	cacagagtgc	ttcgaggagt	tggttacatt	7140
catthttctt	aaatggaatg	ctgcttatta	thtttctthta	aattgtgcat	ttaagtaaca	7200
cacttcttgt	ttatatgata	tatgtataaa	tgtaatthtt	thtttttgaga	tgaggtthtcg	7260
ctcttgthtc	ccaggctgga	gtgcaatggc	actatctthg	ctcactgcaa	cctccacttc	7320
ctgggttcaa	gtgattctcc	tgctcagcc	tcccagtag	ctgggattac	aggcatgcgc	7380
caccatgccc	ggctagthtt	gtatthttta	tagagaaagg	gtthtctccat	gttggtcagg	7440
ctggtctcga	actcccagacc	tcaggtgatc	cgctgcctt	ggcctcccaa	agtgttgga	7500
ttacaggtgt	gagccaccgt	gccaggccct	gaatcagatt	taaaagaggg	catttcatta	7560
aaaaaaatth	thttgtgtht	gctthtgaga	cagagtctcg	ctctgtcgcc	caggctgcag	7620
tgcatthgga	tgatctthggc	tcaccgcggc	ctcagctcc	caggttcaa	tgattctcct	7680
gcctcagcct	cgactagtht	gagattacag	gaatgcacca	ccaccacagg	aatgcacctg	7740
tctaacttht	gtatthtttag	tatagaggga	gtthtgccat	gttagccagg	ctgctcttga	7800
actcctgacc	tccggtgatc	tgctcgctc	ggctcccaa	gttctgggat	tacaggcgtg	7860
agccaccaca	cccggccgaa	agagggcatt	tcagaatgag	ggtctagcat	aagcacagag	7920
aagggggagc	aataagaggg	aaacagggag	taggtcattt	ttgcaatagc	ctgtgacatt	7980

tgtagggcag	tactggcggg	gaataattaa	gtaaaattgg	ctggtgctgt	ggctcatgcc	8040
tgtaatccca	gcacttttggg	aggccgaggc	gggcaggttg	cttgagccca	ggaattcaag	8100
accaacctgg	gaaacatagc	aagaccctgt	ctcaacaaaa	aagtaaaaaa	attagctggg	8160
ggcgcgatgg	ggtggctcat	gcctgtaatc	ccaacacttt	ggaaggctga	ggcaggcgga	8220
ttgcttgagc	ccaggagttg	gagaccagcc	tgggcaacat	ggtgaaaccc	tggctctata	8280
aagaatacaa	aaattagtcg	ggcccagtg	cgtgtgcctg	tgatcccagc	tactcgggag	8340
gctgaggtgg	aaggatcacc	tgagccaggg	aggtggaggt	tgagtgaggt	catgttgttt	8400
gcgccactgc	actccagcct	gggcaatgga	gtgaaaccct	gtccaaaaaa	taaaaaata	8460
aagctgtggc	agaatgtgga	gattcttgga	agctggaagc	tctcatgggg	catttggaag	8520
cctcacattg	taaataacgg	agtcttttta	tcagtttggc	ttccttagtt	ttaggaaaca	8580
agaaataatt	atggctaact	caagtaaaaa	gagaaagaga	agagaaaaaa	gacgtggaga	8640
tagagagaga	gggagagaga	ggaaaagacg	aaaggaagga	aggggggaaag	gagagaggaa	8700
gagagaaaca	gagaaacaga	ctgattagtg	tattggatag	attacataac	caagtacca	8760
gtcaggaacc	cagcagctct	gggggagctc	aatgtgatgc	attgataaac	ccgctcttaa	8820
gagcactcgt	ttccagttac	tttctattcg	gtgggtctcc	agccaagatt	ccaggtccca	8880
ggagaatctg	actgacctag	tgtttgcttc	cgcttttgcg	gtctgggttc	tgtgcttgca	8940
gctcattaga	atacagggag	cagagacaag	caggtagttt	cccaaaggaa	gggatgctga	9000
gtagattaaa	aaaaaagtgt	agattcttca	gtaaactatg	ggatggtaac	tatgcaaac	9060
ctaagatttc	ccttattcaa	ataaattatc	tttcatatta	gacatctaaa	tatgcactaa	9120
tttagttaaa	cccctgggtt	agttgatctc	atcacactga	gctaacattt	ttgttgctgt	9180
tgtttgcagt	gacctgaagt	ttcttatctt	cacaattgct	ttcctctcaa	ataattccca	9240
gatttttaa	ttttatttta	ttttttctgg	agacggagtc	tcgctctgtc	gcccaggctg	9300
gagtgcagt	gcgcgatctc	agctcacttg	cagcctctgc	ctcccaggtt	caagcgattc	9360
tccggcctta	gccttccaac	cagctgggac	tacaggcgcg	cgccccacg	cccggcta	9420
ttaattccca	gattgatatc	cattgcttct	gagatgggcc	aattatcctt	cggagaagac	9480
ttaggtcgcc	tggcagaaaa	agatgaaaga	aatctaagaa	aacgacgaca	ctgagagagg	9540
agcctagcga	accagcagag	cgaccccaag	ccgcaattcc	cccttccgtg	gatcgattac	9600
gaaggcttcc	tggcaggagc	tctccagggc	tgccgacgtg	agccgaactg	cacattggga	9660
actgtagttg	agtgggaaag	ccgagaggcg	ggggccgcac	acgcgtacag	gggccccggt	9720
caacaaagac	gcgccgtg	cgcgcgcgcc	ggagaaaaac	acgggaagac	gtgcgcgtgc	9780

gcggccgccc	ctgggagcca	ccaggcgggc	gagaggacag	cgacaggaag	ggagggggccc	9840
gagccaccga	ctgcaggagg	agaaggggtt	gtgctcctgg	ccgaggaagg	agaaaggggc	9900
ggggccggcg	ggcagcgtgc	ggcagtgcgg	agctcctgag	accggcgggc	acacgggggt	9960
ctgtggcccc	cgccgtagca	gtggctgccc	ccgtcgcttg	gttcccgtcg	gtctgcggga	10020
ggcgggttat	ggcggcgggc	gcagtgcgag	ctgtgaatga	attctccggg	tggacgaggg	10080
aagaagaaag	gctccggcgg	cgccagcaac	ccggtgcctc	ccaggcctcc	gcccccttgc	10140
ctggcccccg	ccccctccgc	cgccggggcg	gccccctccg	ccgagtcgcc	gcataagcgg	10200
aacctgtact	atttctccta	cccgtgtgtt	gtaggcttcg	cgctgctgcg	tttggtcgcc	10260
ttccacctgg	ggctcctctt	cgtgtggctc	tgccagcgct	tctcccgcgc	cctcatggca	10320
gccaagagga	gctccggggc	cgcgccagca	cctgcctcgg	cctcggcccc	ggcgccgggt	10380
ccggggcgcg	aggccgagcg	cgtccgagtc	ttccacaaac	aggccttcga	gtacatctcc	10440
attgccctgc	gcatcgatga	ggatgagaaa	ggtaactagg	gggctggggg	agggggcggc	10500
ggcgccggga	agaaggcggt	ggggtcgccc	ggggagggca	acacctgcgt	cccttttctg	10560
cgggagggga	cgggtgcacc	ccggaattga	tatgccccgg	gagactgctt	tcccgtaggt	10620
cggagcctca	tcttctagta	ttcttaaaac	ctctcccctt	tcagggcact	gtagctgtcg	10680
actttgtttc	agacaccagc	cttccccccac	acttctgcat	gacccaggtc	actatgagac	10740
accagacgt	gttgatgaca	gtgacatttg	tcctagagtg	accacactga	tcctttctag	10800
cactgtgaag	agtgtgcagc	ttcctctgaa	ccaaggtttc	caaaaggttt	tgatattgaa	10860
gaagcagtgc	cgccttactg	gcttttaaat	aaagcagagt	attgtagtgt	cagaaaaaaa	10920
gaacaaatgg	tgacaatttt	gaaagaaata	gctgcatatg	actgcagttg	aatttgcata	10980
attttaatca	agataatcat	tttatcagta	caacgattcc	tgaatacttt	ttcaatgaag	11040
ttatatattag	cataaaactt	ttccccctgt	tgctttgatt	ttaattaaaa	cagtattcca	11100
aagtagcctt	taattttcaa	gttgaaatgt	ttgatgaatg	gattgcgtaa	acttaaacad	11160
accactttac	agtaaaacct	aaaacaacta	tgtatgtttc	tgaatgaaag	caaggatact	11220
acatctttcg	ggtttctttt	aagctaactt	tttttttttt	ttgagacgga	atctcgccct	11280
gtcgcccaag	ctggagtgc	atgggtgcagt	ctcggctcac	tgcaacctcc	gactccctgg	11340
ttccagcgat	tctcctgcct	cagcctcccg	agtagctggg	attacaggca	cgcaccacca	11400
cgcctggcta	atttttgtat	ttttagtaga	gtagggattt	caccatgttg	gccaggatgg	11460
tctccatctc	ctgacctcgt	gatcgggtcg	cctcggcctc	ccaaagtgct	ggtaatacag	11520
gcatgagcca	ccgcgcctgg	ccttaagcta	acatttttta	ttatatgtgc	caggcattgt	11580

gctattagct	ttgcatgtgt	tatctttttt	cttttttaaaa	aaaatagcaa	ccatcccaga	11640
ccatgaaaag	tgttatttta	tcctcacaat	aactttgtga	gatgaaggta	ttattggtat	11700
cagcatttta	gagatgaaga	aaatgaggcc	caaaagataa	aggaggttat	tccaaacctg	11760
tgctaataat	gaaatttctt	ttatggaaaa	taagtgaat	taggagaagt	ctaactttta	11820
caattctccc	cttttatcct	tacttccagt	atgctgagat	cttgcttctc	cctctgccaa	11880
aaacacccac	ttttctacca	cacctcaatt	agatactcac	ttgcattgtc	cattagttaa	11940
aacagaaaca	atctgcactt	cattcatagt	gtctgtctct	actgccaaact	ccaaaaactg	12000
ttctagaatt	tcaatcttgt	tgaaacctgt	ttcctttgtg	gggcctggga	gtgggaagtg	12060
ggatataagg	aaagaggcag	ttattcattg	ttttggacag	taaggaaaga	gtgacggtta	12120
agagaggtca	aggagggtgt	taacatttta	gaatactatg	tgttttaga	aggaaatttt	12180
ttgttaactg	ctccattctt	tttttttttt	ttttttgaga	tggagtctcc	ctgtgtcacc	12240
cagactggag	tgcagtggcg	tgatctcagc	ttactgcaac	ctctgcctcc	cagacgcagg	12300
cgattctctt	gcctcagtct	cccgagtagc	tggaaccaca	ggtgcgccct	atcacggctg	12360
gctaattttt	tgtattttta	gtagagatgg	tgtttcacca	tgttggccag	actggtcacg	12420
aactcctgac	ctcaggtgat	tcgcctgcct	cgaccttcca	agtgccttga	ttacaggtgt	12480
cagccactgt	gcccaccag	caagctccat	tctttatcac	ctcttaagaa	catccaggat	12540
cccttgggga	gaattaaaac	ggttgcaaag	ttttagaata	gaggaacatg	tttaagcgta	12600
gatcattttt	tgggccagtc	atggtggctc	acgcgtgtaa	tcccaacact	ttgggaggcc	12660
ggggcggtg	gatcatgagg	tcaagagttt	gagaccatcc	tggccaacat	ggtgaaatcc	12720
catctctact	aaaaatacaa	aaattagctg	ggcatggtga	cacatgcctg	tagtcccagc	12780
tactcaggag	gctgaggcaa	gagaagtgtc	tgaacctgga	aggtggaggt	tgcagtgagc	12840
cgagatcgtg	ccactacact	ccagcctggg	cggcagagca	agactccgtc	ttgaaaaaaaa	12900
aaagtaggtc	atttttggct	gggcacggtg	gctcatgcct	gtaattccag	cactttggga	12960
ggctgaggtg	ggtggattgc	ttgagcccag	gagtttgaga	ccagcctggg	caacatagtg	13020
aaaccctgtc	tttgtgaaaa	atacaaagat	tagctaggcg	cagtggcaaa	tgctgttagt	13080
cccagctact	tggggggctg	aggtaggagg	atcacttgag	ctcaggttgt	tcaggctgca	13140
atgagctgag	atcgtgcaac	tgcaactcga	actgggtgac	aggagtaaaa	ctgtctcaaa	13200
aaaaaaaaaa	aaaaaaacca	aaaaaaaaaa	aactgtttta	attgttttat	ttaggaagag	13260
aaagtcagaa	catgcaagga	aatttttttt	atgtgtttat	ttttgagacg	gagtctcgct	13320
cagttgccca	ggctggagtg	caatggtatg	atcttggtc	actgcaacct	ctgcctcccg	13380

gattcaagcg	attctcctgc	ctcagcctcc	tgagcagctg	ggattacagg	tgtatgccac	13440
cacgcccagc	taatttttgt	gttttttagta	gagatgggggt	tccaccatgt	tggccaggct	13500
ggtttcgagc	tcctgacctc	aagtgaaccg	ccctccttgg	cctcccaaag	tgctgggatt	13560
acaggtgtga	gccgcggtga	ccgaccacaa	ggaaatttta	gttaacactg	ttggttgatg	13620
ggagttggga	ggtaggataa	aaggagaaat	taaggaaaac	ctaggcatga	aaaataaaag	13680
accctgagct	cttagatttg	aagaaatagc	agttccatgt	gaggaataag	tggaagaaat	13740
agaattcaga	cctcaggtct	caggctggtg	acttaaactc	ttcagtatca	catatatgaa	13800
tatatctatt	atacttaaata	cctcctaaac	atttttatgt	ttcagttgga	tatattaaat	13860
atataaaaaat	aattatttaa	tttatttttag	agacagtgtt	tcactctctt	acccaagctg	13920
gagtggaaatg	tgatcatagc	tcactgcagc	ctcaaggctc	attcctaggc	ttaagtgatc	13980
ctcttttttt	tttttttgct	agagatagga	tcttgttatg	ttagccaggc	tgagagaaatt	14040
tcaacaatat	tttgaacaat	aaaaaaaaaa	taaattaggt	tttattgtaa	agtgggatgt	14100
ttaagtttac	gccattctcc	tgcgtaatgc	atatttcata	ctcttcctac	tgataatggt	14160
tctgttcaca	actttttctt	tatattttga	tttcttttct	cttttctttt	tttttttttt	14220
tgagacagag	cctcgctttg	tcgcccaggc	tagagtgcag	tggcgcgatc	tcggctcact	14280
gctagctccg	cctcccgggt	tcacgccatt	ctcctgcctc	agcttcccga	gtagctggga	14340
ctacaggcgc	ccgccaccat	gccagctaa	ttttttgtat	ttttagtaga	gacgggggtt	14400
cactgtgtta	gccaggatgg	tctcgatctc	ctgacctcgt	gatccacccg	cctcggcctc	14460
ccaaagtgtg	gggattacag	gcgtgagcca	ccgcgcccgg	cacatatttt	gatttctaata	14520
gtggacatca	gaatgggctg	ttgacagctc	tttttttttt	ttttgaggca	gagtctcgct	14580
tcttcaccca	ggctggagtg	caatggcgcg	atctcagctc	actgcaacct	ccacctccca	14640
ggttcaagcg	attctcctgc	ctcagcctcc	ccagtaactg	ggattacagg	catgtgccac	14700
cccacaccca	gccaattttt	gtatttttag	tagagacagc	atttcaccat	gttggccagg	14760
ctgttctcga	actcctggcc	tcaagtgatc	cactcgcctt	ggcctcacia	actgctagga	14820
ttacaggtgt	gagccaccgt	gccagcctt	gacaggtctt	tagtttgatt	ttagttcaac	14880
aactgatgcc	gtaatatgcc	aaattaaatt	agttcagact	gaaacggatt	acttaaagat	14940
tcattttcct	ttaaaaatga	agtaaaactt	tagccggatg	tggtggcggg	tgtgtgtaat	15000
cccagctact	cgggaggggtg	aggcaggaga	atcgcttgaa	cccaggaggt	ggaggttgca	15060
gtgagctgag	atctcaccca	ttgcactcct	gcctgggtga	gaagagtgag	actccatcaa	15120
aaaaaaaaaa	aaaggctggg	catggtggct	caagccgggt	gcagtggccc	atgcctgtta	15180

tcccagcact	ttgggaggca	gaggcaggcg	gatcacttgt	agtcagaagt	tcgcgaccag	15240
cctggccaac	atattagtga	aaccccgctc	ctactaaaaa	tacaaaaatt	agctgggcat	15300
agtgggtcac	gcctgtagtc	ccgggtactt	gggacgctga	cgcaggacaa	ttgcttgaac	15360
tccagaggtg	gagattgcgg	tgggccaaga	tcacgccact	gcacgccagc	ctgggcaaca	15420
gagcaagact	ccatctctca	aaaaaaaaaa	aaaaagaaaa	ctgagtttat	attgtttatg	15480
ttcttatcta	tcttaaattt	tttcttttag	agattgaata	ttttgtact	ttagtcttaa	15540
tccagtgggt	gaaagagagt	gcacttttga	agtctgcctc	ttggctgtcc	ttgacaacac	15600
aaaccttagt	tccaagagaa	tgtaattctt	cctctttctc	agtgcttcaa	aatatatatg	15660
atcaaataga	aacgagttga	ataggcagtc	tcttcaaagg	tttcctaact	ctgtgggttaa	15720
ctatcaagga	gctggtaata	tcatgcactg	ccattccctt	ggcaacatga	cttatctttt	15780
tttttttttt	ttcttttttg	agacagagtc	ttgctctgtc	accaggtctg	gagtgcagtg	15840
gcgagatctc	agctcgctgc	aagccccgcc	tcccacgttc	acgccatttt	cctgcctcag	15900
cctccaagt	agctgggact	acaggcgctt	gccatcacac	ctggctaatt	ttttttgtat	15960
tttttagtag	agacgggggt	tcaccgtggt	agccagtacg	gtctcgatct	cctgtcctcg	16020
tgatccgccc	accttggcct	cccaaagtgt	tgggattaca	ggtgtgagcc	actgcgctg	16080
gcctacttat	cttctaattt	aactgaaaac	caatttat	gattcagtga	aatggcatca	16140
aactgtagta	gtgttaattg	aaatatttgg	taccttgaaa	tgtaaatagc	caaattaaat	16200
ctcattttaa	tgaaatctcg	tgtaaagtgt	ttttatatgg	tgactatggt	tattctgaat	16260
tttattctta	tggcatacta	aaaaaaaaaa	aaattttttt	ttttttttgt	aatggagtct	16320
tgctctgtca	cccacgctgg	agtgcagtgg	cgcaatctcg	gttcaactga	acctccacct	16380
cccggttca	agcgattctc	ctgcctcagt	ctcccagata	gctgggtacta	caggcgtgca	16440
ccactatgcc	tggctaatat	ttttgtat	ttaggagaga	cagggtttca	ccatgttggt	16500
caggctgata	tcgaactcct	gacctgaagt	gatccgcctg	cctcagcctc	ccaaagtgct	16560
gggattacag	gtatgaccca	ctgcacccaa	cccatactca	aatttgacac	tgaattttca	16620
taaaggcctt	aatttatctg	aaaccaaact	atttcaaaag	aggaatagca	cagcaaattc	16680
tgttgactta	atgagaggat	atgtgaagtc	tatttattaa	agcaaatatt	aattggaggc	16740
cagttaattt	gtacagctct	gcattttaga	tatttgagaa	atatttat	cctctccagt	16800
gagatgtgtt	aaaacattag	ttatgtgatt	aacaaatatg	tgtacatacg	tatatatgta	16860
catacacatt	ttgagacagg	gccttgggtc	gttggttcagg	ctggagtgca	atgacaccat	16920
cttagctcat	tgaagcttca	gcctcgcagg	ctcaatcgat	ccaccacact	cagcctccct	16980

agtagctggg	tctacagaca	tataccacca	tgcttggctg	atTTTTtaat	TTTTgtaga	17040
gatggtgatc	ttgccctggt	gactaagttg	gtaaatattt	taattgttga	actttcttgg	17100
aagactgaaa	acctgtgata	gccattttat	ataaggagaa	gctgaagttc	aaagagtaga	17160
ctcatagcag	aaacaaaaat	agaatttaag	tgaatggact	caaaatattg	tactttttac	17220
tttatactgc	aggtttttat	gttgtaatgc	tggtaatgag	ctccttggaa	tatttggagg	17280
aaaagagaag	gttgtaataa	tggttctttg	gatttactag	aacatatcat	gttctgcatg	17340
gctcctgtgg	gtagacaagc	cggaaatctc	ctgggtaaca	caatgggtga	ggttctctag	17400
gtgacgtttg	atttctcaag	tacataggac	taaacagaaa	aggcctagta	tgttatatga	17460
atgagagatc	aagtttctca	ggatattcta	gggctaaagg	atcaggcatc	gaagacagaa	17520
attgtctaaa	taaaattttt	tctattcata	gttttaaagg	gctaaagggg	caggcattga	17580
agacagaaat	tgtctaagta	aaatattttt	ctgttcatag	ttttaagctg	tgtatatgtg	17640
catgtgtgta	tttaaaataa	cttcgtcaat	gaaaagagtc	aaactctgta	aaatatttga	17700
agagatttat	tctgagccaa	atatgagtga	ccagtggccc	atgacacagc	cccagtagat	17760
actaagaaca	tctgtccaag	gtggtcaggc	tatagcttga	ttttatacac	tttagggaga	17820
cataagacgt	cagttaaaca	tgtaagatgt	acattgggtc	catctggaaa	ggcaggaaaa	17880
ctagaagttg	gggaggcttc	caggtcgtag	gcagattcaa	agattttctg	attggcaatt	17940
ggtcaaaaga	gcttatctaa	agtcctggaa	tccatagaag	ggagtgtctg	gtttaaaata	18000
ataggttgta	gctaccaagg	tttttattgt	acagatgaag	cctccaggta	gcaggcttca	18060
gagagaatac	attgtaaatg	tttcttatga	gactttaaaa	ggtggcagac	tcttaagtta	18120
attttctcct	ggttcaggta	aaagacttgg	aaagggaaag	gattctctac	agaacgtaaa	18180
ttttccccac	aagagaaagc	tttgcagggc	catttcagaa	tatgtcaaag	aaatataatt	18240
tagggtaaaa	tacttcaatt	tgttttattt	atttatattat	tttttgagac	agagtctcgc	18300
tctgttgccc	aggctggagt	gcagtggcat	gatctcggct	cactgcaagc	tccgcctcct	18360
gtgttcacgc	cgttctcctg	cctcagcctc	ctgagtagct	gggactacag	gcgctggcca	18420
ccatgccag	ctattttttt	tgtattttta	gtagagacgg	gatttcacca	tgttggccag	18480
gatggtctcg	atctcttgac	ctcgtgattc	acccgcctcg	gcctcccaaa	gtgctgggat	18540
tacaggcgtg	agccactgtg	cccggcctca	aaatacttca	atttctttca	tggcctgcta	18600
tctgacgtga	tgctgtacta	gagtcaggct	gggaatttgg	cgtcttattg	ctacaaaaca	18660
tcttaatatc	tctgttttaa	tgttaatgct	gatcagttgt	ccctgaattc	caaagggag	18720
agggtatatg	agtcatgtcc	aacccccact	tctcattatg	gcctgaacta	gttttttagg	18780

ttaacttttg	aatgcctttg	gcaaggggag	ggtccatgag	tcagttgggg	gtcttagagt	18840
tttatttttt	gcttaccggt	ttataaaaaag	ttaatgaaaa	ttatcatatt	tcataattct	18900
atataattca	atattgtact	ttattttaaaa	ctcacgtata	aaatagctgt	ccatatctgt	18960
tttcagaaga	tgaggatgga	gggtagaaat	cagaagtgtc	agatttggtg	attttcttac	19020
actgctgaaa	acctatacct	accactttga	aaggattaat	ttcagacttg	ctttcttttg	19080
gcctaattgat	tctactttga	agtttctctg	attaaactaa	ggaataaatc	tgataaatgg	19140
acattcagat	gataccatac	ttttccaaaa	gataacattg	cttttgatta	catatgcaat	19200
aaacatttca	cattttttct	caagattatt	tactggcatc	tgcaccaaag	acacaaaaaa	19260
gcagccactg	ttaaggactt	tatcccttgt	tctgttttta	gctggtttgt	tggtgtattt	19320
ttcctcatgt	tgaatacagt	taaaccctat	taaactggat	tcccatatt	actgttagtt	19380
gtcctgatga	caattaagga	tagttagtaa	atggatattg	aatcatttta	tttttttagt	19440
agtgatggcg	tttcgctgtg	ttggccaggt	tggtctcgaa	ctcctggcct	caagtgatct	19500
gcctgtgttg	gccctcaaaa	gtgttaggat	tacaggcgtg	agccactgca	cctggctgga	19560
tattgaatta	ttgaagaagg	atgttcccta	acacttcctg	cctcttttcc	gctgtcttac	19620
tctctctgtt	ccaccagta	gtagttgggt	ttttgatgtc	tggcaaagta	tagattgtct	19680
actcttctat	attagttttc	ttcactcttt	tttttaaaaa	tgttttttaa	gcaagataga	19740
gacagggctc	tgccgtattg	cccaggctgg	tcttgaactc	ctgagctcaa	gcaatcctcc	19800
cacctcacc	cataaagtgc	ggagattaca	ggcatgagcc	accatgcctg	gccagttttc	19860
ttcactcttg	acctataata	gtcctgcaaa	gccagtgaag	ctgttaatat	gctgacgtag	19920
cctttttttc	tctcgtttgt	gaattattaa	caatcgctga	tcttacacat	catatacaat	19980
aaaacatctt	tgaacactgt	cacatctcag	cagctcattc	tggttaatga	ggaaagaaaa	20040
atgtcaaaat	ctgtgatttt	cttagagggt	attaaatgtt	ttacagctat	gtagatattc	20100
tgtagacttg	ctgtacttac	ttacataact	tttctgctct	tctgcagagg	gagagaatta	20160
acttcataag	tgggctttgt	caatgccttc	cagtcagcga	tctccaggaa	caaacttatc	20220
tatggttgag	caagtggggg	ttattaccca	ctgtagccag	ggagaacaca	catatgtaat	20280
aacatggta	tgatcatagta	aagggtgtta	ggaagatagg	atttgggctc	gtttatttgg	20340
tgattttaag	gagggtttaa	agaagcaggg	ttttgctctg	atttggatgc	taccaggagc	20400
tgogaataat	tctatgaata	aatatcataa	aacctatcta	gaagaaaaga	ctagagttag	20460
gtctaaagct	gtagtagtta	aaaagcaaa	gtcactcctt	atctggaaaa	ggggaatatt	20520
tggtattttca	tggttttagac	agtgttcagt	gttcatgttt	tgcctgtgtt	tagacataat	20580

tgtagagtgg	tcttggtcca	ccgtgggtcac	agagtggcat	tatttcacgc	tgatattctg	20640
taagtgccag	gtcaaatacc	aagggttagc	tgatagtagt	aggccagctc	ctggatgtaa	20700
taagctgttt	ttctctttct	cagctttttt	ttttgttggt	ttttttttaa	aatagagtct	20760
ggctctgtca	cccaggctgg	agttcagtgg	gggcaatctc	agctcactgc	agcctctgcc	20820
tcctggattc	aagcgactct	cctgcctcag	cctcccaact	agctggaatt	acaggcgcaa	20880
gtcaccacac	ccggctaatt	tttgatattt	tagtagagac	ggggtttcac	catgttgagc	20940
aggctggtct	cgaactccgg	acctcaagca	atccatctgc	ctcagcctcc	caaagtgctg	21000
agattacagg	catgagtcac	tgcacctggc	ctgtttctta	gatttgaggg	tcaactttta	21060
cccttttcga	actgtggggc	ttcttatgga	aattgacatt	taagtctga	ccatatagga	21120
tcttgggcaa	gttacttaac	tatgtatgct	gagatgtttt	tttaaagtgt	tagtgcttgc	21180
acataatagg	tactcagtaa	atgagaacta	ttattataaa	atcaatagta	cttttaagat	21240
tacagctaga	ggttatgtta	gaacattggt	aactcttcgt	actagtttct	gttacacttt	21300
ttaggatgca	gtagcagcat	aaacgagata	tggggaagaa	cgagaggat	ttaaacagtg	21360
atataggccg	ggcgtggtgg	ctcacgcttg	taatcccagc	actttgggag	gctgagggtg	21420
gcagattacg	aggtcaggag	attgagacca	tcctggctaa	cacggtgaaa	ccccatctct	21480
actaaaaata	caaaaacaaa	attagccggg	cgtggtagtg	ggcgccgtga	atcccagcta	21540
tttgggaggc	tgaggcagga	gagtggcgag	aaccaggag	acagagcttg	cagtgagccg	21600
agatcgcgcc	actgcacccc	agcctggggc	acagagcgag	actctgtctc	aaaaaaaaaa	21660
aaaaaaaaaa	attaaaacaa	aatattttgt	gttaattgtg	atgacaaaaa	aaaaaaaaaga	21720
gatggaagtc	tctccctaac	ctcactctc	atttagtgct	atggcttttt	tctttttttt	21780
tttttttttt	ttgagacaaa	gcctcactct	gtcaccagag	ctggactgca	gtggtgcaat	21840
ctcagctcac	tgcaggctct	gcctcccaga	ttcaagcaat	tctcctgcct	cagcctcaca	21900
agtagatggg	actacaggca	catggcacca	tgcccagcta	attttttggtg	tagtttttag	21960
tagagacagg	gtttcactat	gttgccagag	ctggtctcga	actcctgagc	tcaagtgatc	22020
ccccctcctt	ggcctcccaa	agtgtctggg	ttacaggcgt	gagccctgct	cccagactcc	22080
tggctttttt	tttttttttt	aatgaaaaat	tcaaaatgct	cttttttttt	tttttttttt	22140
tttgagacag	agtctcgctc	tgtcgcccag	gctgaagtgc	agtgggtgtga	tctcggtcga	22200
ctgcaacctc	tgcctcccag	gttgaagtga	ttctcctgcc	tcagcctctc	gaacagctga	22260
gattacaggt	acgcgccacc	acgcctggct	aatttttgta	tttttaatag	agatgggggt	22320
ttgtcatgtt	gtccaggcta	gccttgaact	cctcgccctg	tgtgatccac	ctgccttggc	22380

ctcccaaagt	gctaggatta	taggcgtgag	tcgctgcacc	tggccacaga	aatTTTTtga	22440
agaagataaa	taaggtgaca	tttttaagg	tcaaagaaaa	tgtcaaaaac	tagaatgatg	22500
tctttacata	gggtttaaaa	ctttccaaat	taacagggaa	aataattctt	taccttgaaa	22560
ataaatgttt	gctagtgaag	gcaaatacaa	tctttttact	aaatgtttta	ttaaatTTTT	22620
ttttcttgta	gacacagggt	cctactgtat	taccagggt	ggtcttgaat	gcctggcctc	22680
aagcaatcct	cctacctcag	tcttccaaag	tgctgggact	acagacatga	gccatcacac	22740
tgggtctttt	ttaccaaat	atagtagaaa	gcactttttc	tctaattggtg	aactatgaga	22800
gaattaatca	ggggctatta	gtaattcatc	cctgaattaa	tcagtgatta	taatgctttg	22860
tggtccatgt	agtttgctgg	ggattaacac	accatgaaag	tctaccagga	gatttttttt	22920
tttctttgag	aacaggggcc	ataatcagta	gtccttaaat	gaaatggact	attcccattt	22980
cattatatgt	tgctagggt	ggactcgagc	tccttggctt	aagtgatctt	cccacttcag	23040
actctcaagc	agctgggact	ataggtgtgt	gccaccatac	ctggcttaat	tgagtgtttg	23100
ttttttgttt	ttttttctga	gacagagtct	cgttgtgtcg	cccaggctag	agtgcagtgg	23160
cacaatctcg	gctcactgca	acctctgttt	cccaggttca	aacgattctc	ctgcctcagc	23220
ctcccaagta	gctgggatta	caggcgcttg	ccaccatgcc	tggctaattt	ttgtattttt	23280
agtagagatg	gggtttcacc	atcttggcca	ggctgatctc	gaactcttga	cctcatgatc	23340
cacacacctt	ggcctcccaa	agtgtctggga	ttataggcgt	gagccaccgc	gcccggtttt	23400
aattgagatt	tttagatatc	tattactctg	ctaattttgt	cacttgcaag	ttgccatcag	23460
aaaattgtag	gaaaatggat	atatttgttc	cttggaatgg	tttgtgtgag	aatacttaag	23520
gattaaatag	ataagtaaaa	ctggtgggct	ttatataaca	tagatgagca	aatgtcagga	23580
acatacaact	gtgcacacag	ttcaggagaa	ggaggattta	agttaatcaa	caaatttact	23640
aagtataata	aagatactaa	aagtagtggt	tccataccac	tttattactt	aaagtatcat	23700
catataccct	attttatatg	atttttgcca	caagtcagag	ttaggtaaaa	gaaataactg	23760
cttttcaggt	aaggagtttg	acgcccagac	agattaactg	acttttccaa	aatcatattg	23820
ctattaaatg	gtggaacaag	gacttaaatac	tttgccttct	aactcacata	cttgcaaaca	23880
catatcctct	cactctaccc	caagctaccc	atgttttgac	ccttcttgtg	gcaatctggg	23940
tctcactaat	atttgaaaga	aaacgtacag	tagataattt	gcaagttaat	ctgttacgca	24000
tatctcttac	ctctatttaa	agatgaatat	cagcatttct	gttggtttcta	cagtaacata	24060
ctaaaaaata	atgcagtcca	ggtgcagtgg	ctcacacgtg	taatcctagc	acttttgga	24120
gttgaggcag	gaggatcact	tgaagccagg	agttcgagac	tagcctgggt	atgcaagacc	24180

ccatttctgt	tttttttttt	tgtttgtttt	gttttgtttt	gttttgtttt	tttgagacgg	24240
agtctcgctc	tgtcgcccag	gctggagtg	agtggcgga	tctcggtca	ctgcaagctc	24300
cgcctcccgg	gttcacgcca	ttctcctgcc	tcagcctccc	aagtagctgg	gactacaggc	24360
gcccgcact	acgcccggct	aattttttgt	atttttagta	gagacgggg	ttcaccgttt	24420
tagccgggat	ggctctgatc	tcctgacctc	gtgatcccca	tttctttttc	actgcaacct	24480
ttgcctccca	ggttcaagt	attgtcgtgc	ctcagcctcc	caagtagctt	gggattacag	24540
gcatgtgcca	ccatgcctgg	ctaatttttg	tatttttttg	tagagatgga	gtttctggcc	24600
aagctggcca	ggctggtctc	gaacgcctag	cctcaagtga	tctaccctcc	ttagcctccc	24660
aaagtgctgg	gattacaggt	gtgagccact	gtgcccggcc	ccatttctac	aacaattaa	24720
aaatattagc	ccagtgtagt	ggtgcatgtc	tctagtccca	gctactcaga	aggctgaagt	24780
gaaaggattg	cttgagccca	gaatttcaag	gctacagtga	gctatgataa	tggcattgca	24840
ctccagcttg	ggtgacagag	tgagaccctg	tctctaaaaa	atgaagtaaa	atagtgcaca	24900
agtatagaac	ttgaaaatct	tccttaacct	taccataagg	gaaatgatta	ctaataagtt	24960
tcttaacttt	ttgtacttac	ataaacataa	atattcatca	gagaaaaaaa	tatgcaaaac	25020
aatttgcaat	ctttttcact	taccatattt	tggaattttt	ttcatttcaa	tatatttgat	25080
cttccttg	ttttttcagt	ttgtttgttt	ttgtcaccca	ggttgagtg	aagtggtaag	25140
aacatggctt	attgcagcct	caatctcctg	ctcccattca	gcccctcaag	tagctgggac	25200
tacaggtaca	tgtcaccacg	cccggcta	ttttattttt	attttggtag	agatgggggt	25260
tcaccatgtt	gcccaggctg	atcttgaatt	cctgggctca	agtgacccgc	ccacctcagc	25320
ctccgaagt	gttgagagaa	caggtgtgaa	ccaccatgct	ccacctctta	gtctttacaa	25380
tctgcaaaac	ctcataagtg	gctaatagag	gaatatagta	aagcaaaggg	ggatatcact	25440
gattagaact	gtgttttttag	gctgggtgcg	gtggctcacg	cctgtaatcg	caacattttg	25500
ggaggctaag	tgggagtatc	acttgagccc	gggagttcaa	gaccagcctg	ggcaatatag	25560
tgagagaccc	tgtcttttaga	aaaaaattaa	ccaggtgtgg	tgggtgcacac	ctgtgggtccc	25620
agctattcaa	gaggctgagg	tgggagaatc	gcttaagcct	aggaggcgga	ggttgcatg	25680
agatcatacc	actgctctct	agcatgggtg	acagagcgag	accagtccta	aaaaaaaaag	25740
tattttttcg	tttttttcca	actcatgtac	acccgccacc	ccaccctgc	tttttttttt	25800
tctgacattg	ggtcttgctc	tgtcaccacg	gctagagtgc	agtagcacia	tcaactcact	25860
gcagcctcct	cttcctggac	tcaagcagtc	ctaccacctc	agcctcccaa	gtagctggga	25920
ccacaggtgt	gcaccatcat	gcctggctaa	tttttgtact	ttttgtagag	atagggtttc	25980

accatgtggt	tcaccatctc	taactcctgg	gctcagtcag	tccacttttg	cctcggcatg	26040
agccactgtg	tgcagcccac	gtttttttatt	aatggatatt	tggattgttt	ccatctattg	26100
tgaataatgt	ggctatgaac	attggtctaa	atatctgttt	aagtcccggc	tttcaatact	26160
tttgatata	tacctaggag	tagaattact	gaattatatg	gtaactttct	gtttaaattt	26220
ttgaactgcc	aacctgtttt	ccataggggc	tgcaccattt	tgcattccca	ccagcagtgt	26280
acaagggttc	cagttttctcc	acatttgтта	tttttcattt	tttaaaataa	tagtcatcct	26340
aaagggtatg	aagtggatc	tcgttgtgat	tttgatttgc	atgtattttt	ctaagtactt	26400
atgatgctca	gcattttgtc	atgtacttat	gtaccatttg	tgtatcttct	ttggaaaaat	26460
gtctattaat	gttcttttcc	catttttttaa	ttgggttggt	tttatgttta	tcaattttgt	26520
aaacatttta	agctctgcag	cataactact	caaccctgtc	acatggtaag	attgaccag	26580
taaaacttta	tgtacaaaaa	taggcagctt	actagattta	atcttagtcc	atagtttgct	26640
aaggcatgca	ttagataatg	tagttacact	attggctaат	aattttaaact	acaagtgggt	26700
gtaagtttct	gccacccaaa	ttctttcttg	atttgatgta	gtctggttgg	ttgaatttga	26760
gtgttatatg	gggtcataga	gttaaaagag	aaatgtctat	gagaaaactag	ggactgttgg	26820
gagctaатgt	taaaggattt	tggaggcctt	ttgtgcactg	gagaccattg	gaagattgga	26880
ttttctgcta	taaatgtatc	taaaggataa	tcagtgtaaг	ttatgggctg	tagtttgcca	26940
accctacat	taagggattg	aaataatttg	aatatgggtt	tcagttcttg	tatggtctgg	27000
ctcagttctt	tttttttttt	tttttttttt	tttttttttg	agatggagtc	tcactctggt	27060
gccaagctg	cagtgccatg	gcacgatctc	agctcgctgc	aacctccgcc	tcctggcttc	27120
aagtgattct	cctgcctcag	cctccagagt	agctaggact	acaggcacgt	gccaccacat	27180
ctggctaatt	tttttttttt	tcaagacaga	gtttcgctct	tgtcacccag	cctggagtgc	27240
aatggcacga	tcttggtctca	ccgcagcctc	cgctcctgg	gttcaagtga	ttctcctgcc	27300
tcagcctccc	aaatagctga	ggcaacaggc	gtgcgccacc	acgcctagct	aatttttctg	27360
tttttagtag	agatgggggt	tcaccatctt	ggccaggctg	gtctcgaact	cctgatctca	27420
tgatccaccc	gcctcagcct	cccaaagtgc	tgggattaca	ggcgtgagcc	acctcacctg	27480
gcgttttttt	gtgtgtgttt	tcagtagaga	tggggcttca	ctgtattagc	cagaatggtc	27540
ttgatctcct	gacctcgtga	ttcacctgcc	tcagcctccc	aaagtgcagg	aattataggc	27600
atgagccacc	acgcctggct	gtggctcagt	tcttaactgt	tcattcatte	agggteccag	27660
ccaaaatctt	ttgccacagc	ttctcctggg	catgctctga	aattcatttt	ttgtctgttt	27720
agttgcatga	gtccactgaa	agctttgctt	attttctcca	ccactttcag	agttcacagt	27780

cacttcaagt	agacgaggtc	aggagatcga	gaccatcctg	gccaacatgg	tgaaacccccg	27840
tctctactaa	aaatacaaaa	cttagctggg	cgtggtagcg	tgtgcctgta	gtcccagcta	27900
ctcaggaggc	tgaggcagga	aaattgcttg	aacccgagag	gcagagggtg	cagtaagccg	27960
agatcatacc	atttcactcc	agcctgggcg	acagagtgag	acactgtcaa	aaaaaaaaaa	28020
aaaaaaaaaa	aggaaagaaa	agtgtcctaa	atgctgggct	catctctgct	ggcttctctc	28080
ctttttatat	ttcgataatt	tttaaaataa	taaccatcct	aatgggcttg	atatggtccc	28140
tcctttctaa	attttggtaa	ttttcattgt	tttgaaact	cactaatccc	ttcaaacaga	28200
tgtttttaaa	actttttttc	tagttctctc	tctttttttt	tttttttttt	tttttgagac	28260
agggtctctc	tcagtgggtgc	gacctccctt	actgcaacct	ctccctcctg	ggttcaagca	28320
gcgattctcc	tgcctcagcc	tcccaaatag	ctgggagtac	aggcaccac	ccccacgccc	28380
agctaattgt	tgtattttta	gtagagattg	ggtttcatca	tgttggccag	gctggtctcg	28440
aactcctgac	cccagggtgat	ccgcctgcct	cagccttcca	aagtgtctgg	attacagcca	28500
tgagccaccg	cacctggctc	tagttctctt	ctaattctac	tatagtaatc	aaaatttttag	28560
tgctgatgtt	ataatccaag	cccaaattgga	tattcttata	cattaaatgt	tggaatatca	28620
tgtttggtta	aaatcagatc	tgctgggcac	ggtaggtcaa	tgctgttat	cccagcactt	28680
tgaggaggccg	aggcagggtg	atcccctgag	atcaggagtt	cgagaccagc	ctgaccaaca	28740
tgagaaaact	ccgtctctac	taaaaataca	aaattagccg	ggcgtgggtg	cacgtgcctt	28800
tcatcccagc	tacttgggag	gctgaggcag	aagaattgct	tgaacctgtg	aggcgagggt	28860
tccagggagc	tgagatcgtg	ccattgtact	ccagcctggg	caacaagagc	gaaattccat	28920
ctcaaaaaaa	aaaaatcaga	tctgtcccta	tggttttatc	ttcactgcat	gtcatatata	28980
aatggaagca	tgtagtatgt	agccttttgt	gtctcgcttc	tctccctttg	cataatattt	29040
ttgagattta	tcccttttcc	tacatatgtc	agtagtttgt	ttctttttta	tgctgaatag	29100
tattccacgt	gtggagatac	cacaatttgt	tactccattc	actaaatgat	ttgggttggt	29160
tcctgttttt	gttattgttg	ttgttggtgt	ttaattaatt	aattaatttt	ttttttgaga	29220
tgaggtcttg	ctctgttgcc	caggctagag	tgagtgggcg	caatcacagc	tcactgcaac	29280
ctctgcctcc	ccggttcaag	tgattctcct	gcctcagcct	cccaagtagc	tgtgattaca	29340
gggtgcccgc	accacaacca	gctaattttt	gtatttttag	tagagatggg	atttcaccat	29400
gttggccagg	ctggtcttga	actcttcctt	ccttggtctc	ccaaaatgct	gggattacag	29460
gtgtgagcca	ccatgcccac	cctcctgttt	ttaaatttta	tgaataaagg	tactgtaaat	29520
tcatgtacag	gtcttagaag	tttgcatttt	tcttgggtta	atatgtagaa	gaagagattg	29580

ttgactcatg	tggtaaatat	atgtttaatg	tcataagaaa	ctaccaaact	gtttttccaa	29640
ctgggtgcc	ttttgtttcc	taccagcaaa	cataatgaaa	gggccattac	cctttgtctc	29700
gtaatatatt	ggatatttta	tgttcttttg	tttttttgcc	attcaaaatt	gggtgcattg	29760
ttgtatcttt	ggattttta	ttgtacttcc	ctaacaacta	atgatgtgga	gcatcttttt	29820
atacgcttac	ttgccaatta	tgtcttaatt	ctttggcaat	acttatattt	gactctgcct	29880
actatatata	tatatatata	tatttttatg	ctatatatag	agtatatata	ttttttatct	29940
atatatatatt	atatataacta	tatatattta	tataacttat	atatatatat	atatatatat	30000
atattttttt	ttttttttga	gatgaagtca	tgctctgttg	accaggctgg	atgcagtggc	30060
atgatcttgg	cttaccgcaa	cctcaacctc	tgctctccag	gttcaagtga	ttctcctgcc	30120
tcagcttcct	gagtagctgg	gactacagac	acatgccacc	atgcccggct	aatttttgta	30180
tttttagtag	aggtggagtt	tcgccatggt	gtccaggctg	gtcttcaact	cctgacctca	30240
gggtgatccac	ccgcttcagc	ctcccaaaag	tgctgggatt	acaggcatga	cccaccgtgc	30300
ccagctggaa	taagcaatct	taaaaagcag	tttgttgatt	tctggtgaat	agagaaaatg	30360
tgacaaacta	tgactcaata	gtttcatttc	ttttttgttt	gtttgtttgt	ttttttgaga	30420
cagagtcttg	ctctgtcacc	caggctggcg	tgagtggtct	cactgcaagc	tctgcttcct	30480
gggttcacgc	cattctcctg	cctcagcctc	cggagtagct	gggactacag	gcgcccggca	30540
ccatgcctgg	ctaatttggt	ttttgtattt	ttagtagaga	cagggtttca	ctgtgttagc	30600
caggatggtc	tcgatctcct	gacctcgtga	tccaccgctc	tcggcctccc	aaagtgtctg	30660
gattacaggc	atgagccacc	gcccttgccc	tactttttatt	tcttaagcat	atatcttaaa	30720
gaaactgcag	gtatgcatta	ggggacaggt	agaaaaatgt	tacagcattg	tttgaaatgt	30780
tgaaaaactg	gaaacaacct	caatatctta	ataatagaaa	aaatgggtga	ggctggctgt	30840
gggtggtcat	gcctgtaatc	ctagcacttt	tggaggctga	ggcgggcgga	tcacctgagg	30900
tcgggagttt	gagtccagcc	tggccaacat	gggtgaaacc	cgtctctact	aaaaatacaa	30960
aaaattaggc	aggtgtggtg	gtgcacacct	gtaatcccag	ctacttggga	ggctgaggca	31020
caagaatcac	ttgagcctgg	gaaatggagg	ttgtagttag	ccgagatcat	gcctactgca	31080
ctccagcctg	ggcaatagag	tgagactccg	tctcaaaaaa	aaaaaaaaaa	aaaaaaaggg	31140
aaggagcatg	ggaatgggaa	gaggcactaa	aaaaggaact	tcttttgtat	ctgttacatt	31200
gtactcattt	tgtttttaaa	attttaagcc	aaatatgaca	ttgttatcac	ttttgttgcc	31260
actcagaata	caagtgtctat	tacactattt	tctataattt	tccctttttc	aaactaaaaa	31320
acaaaaatac	atgtaaacag	gtgttttctt	tatggaaaac	aggaaggtga	gcatccaaat	31380

aatttttagtt	aattttataac	aaactcttaa	taattttcttt	ttctttcttct	tctttttttt	31440
tttttttttt	tgagtcaggg	tcttgctctt	ttgctcaggt	tagagtgcaa	tggtacaatc	31500
tggtacaatc	acatcttagt	gcaacttcta	cctcttcagc	tcaagtgacc	tctctcctca	31560
gtccccaag	tagatgggac	cacagatgtg	tatccccaat	catggttaat	ttttaaattt	31620
tttggtagag	ataaaatctc	accatgttgc	ccaagctggg	cttgagctcc	tgggctcaag	31680
cagtcctcct	gccttggcct	cccaaagtgc	tgggattata	ggcgtgagca	aactgttccc	31740
agtcttcata	aacttttctt	tacatgtcct	tatcaagtac	tttttgagca	cctactgtca	31800
accagccgta	gtatataata	ctgataattc	tataacataa	gaaattgacc	tgtttaaggg	31860
gatgaatatg	gttagttatt	ttcaggggtga	agaaacagaa	gaatcgggga	ggtagtacat	31920
agtcataagg	agacggcatt	tcttagtcac	tttgtgtggt	gtttataata	agacttcatg	31980
tggccaggtg	cggtggctcg	tgagcgcccg	taatcccagc	actttgggag	gccgaggtgg	32040
gaggatcaca	acaaggtcag	gagatcgaga	ctatcctgaa	caacatggtg	aaaccccgtc	32100
tctactaaaa	atacaaaaat	tagctgggtg	tggtggcacg	tgcctgtaac	cccagctact	32160
cgggaggtcg	aggcaggaga	atcgcttgaa	ccagggagtc	ggaggttgca	gtgaaccaag	32220
atcgtgccac	tgcactgcag	cctgggctac	agagagagac	tccatctaaa	taaataaata	32280
aataaataaa	taaatagact	tcgtctgttt	catcaagagt	cattgtatta	tattgatttt	32340
gaaatttcaa	ataataataa	ggaatagtat	ggaaattttt	gtttgattaa	atggggacat	32400
gaggccaggt	gtggtcactc	aggcttgtaa	tcccagcact	tcgggagacc	agggcaggag	32460
gattacttga	gcctaggagt	tcaagaccag	cctgggcaac	atggcaagac	cctatctcta	32520
caaaagaaat	acacaaaaat	tagctaagtg	tggtggcaca	tgcctgtagt	cttagctact	32580
cgggaggcca	aggggaagagg	atggatcact	tgagcccagg	agttggaggg	tgcagtgage	32640
taggattgtg	cccctgcact	ctagtgtggg	tgacagcaag	actctgttta	aaaaaaaaaa	32700
aaaatgaggc	cgggcgcggt	ggctcacctc	tgtaatctga	gcactttgag	aggccaaggc	32760
gggtggatca	cctgagttca	agatcagcct	ggccaacatg	gtgaaaaccc	cgtctttata	32820
cttaaattac	aaagattagc	tgggtgtgat	ggtgcacacc	tgtaacccca	gctactccag	32880
aggctgagac	agcagaatcg	cttgaacccg	ggagggggag	gttgcaatga	gccgagattg	32940
caccactgca	ctccagccta	ggtgacagag	caagactctg	tctcaaaaaa	aaaagaaaaa	33000
aagtctaaga	agttaatttt	cattcagaca	aatgttcaaa	aataatagaa	ataaaacaaa	33060
aaattaattt	tcaaataatgt	gatttgagta	taacgatcac	tttacagaag	ttattctaca	33120
tactttttga	ttatgctgag	tttttaggtc	cttttaaagc	ttaacttgta	tcaggaatta	33180

atgtttttaa	ctagttgttg	ggttcaaaga	aaatagaaat	gtgttattaa	atgccacact	33240
tgtaatctca	tacagcagtt	ctcatgatgc	aatcaattgt	taaataaaaa	cttcctctag	33300
agaggaggta	cccagcaatg	aacacattat	atcagatttg	aatatgagat	taaacaatgc	33360
tttttgtttc	attgttattg	aaatcatatt	tgtatttttc	aaaagtatat	atacttaaaa	33420
attcctaaat	aaaatgtcct	tctgtaatac	agcatttttag	gttataagga	tcaataccat	33480
gggctgataa	gcaaaatagg	actcacggcc	ccaaaatggt	gataaacatc	atgaccatat	33540
tccccataat	ggagttacat	atacacatta	caatgataaa	aatactaaga	agttatatag	33600
taatgtttct	cagacttggt	cacaaccctg	ttttttatgt	attacctctc	aaaagcatga	33660
tttgcaatat	ttagtggtact	cttcatacga	tctatacaaa	taatttttta	ttttaagca	33720
ggacagaagg	agcaagctgt	ggaatggtat	aagaaaggta	ttgaagaact	ggaaaaagga	33780
atagctgtta	tagttacagg	acaaggtaag	attgtatttg	tttatagcca	tcccaaatta	33840
tgatatattc	acatgattgt	ccagatttca	gatctattta	tttatttatt	tatttttctt	33900
tctttctttt	cttttctttt	cttttttttt	tttttttttg	agacagggtc	tttcctctgt	33960
taccaggct	ggagtgcagt	gatgtgatca	tagctcactg	caacctcaa	cttctgggct	34020
caagtgatcc	tcttgcctct	gcttcccatg	tagctgggac	cacaggcgca	cactaccata	34080
cttagcta	tattattatt	attattatta	ttattattac	tagtttttga	gacagagtct	34140
cgctcttttg	cccaggctgt	agtgcaatgg	tgcaatctcg	gctcactgca	atctccgcct	34200
cccggattca	ggctatttat	tctcctgcct	gagactcctg	agtagctggg	attacagggtg	34260
catgccaccg	tgcccagcta	atttttgtat	tttttagtgaa	gatgggggtt	tgccacgttg	34320
gccaggctgg	tttcgaactc	ctgatctcaa	gtgatctgcc	cacctcagcc	tcccaaagtg	34380
gtgggagtac	aggcatgagc	cactgtgggc	cacacttagc	taattaaaaa	aaattttttt	34440
gtagagacag	ggtctcactg	tgtttctgtg	actgatctta	aacttctggc	ctcaagtgat	34500
cctcctgcct	cagctttcca	aagttgtggg	attacaggca	tgagccactg	taccagcca	34560
gattgtattt	atcttaataa	atgtatttg	gccaggcaca	gtggctcatg	cctgtaatcc	34620
cagcactttg	ggaggccgag	gtgggcgaat	cacctgggat	caggagtctg	agaccagcct	34680
gatcaacatg	gagaaaccct	gtctctacta	aaaattcaaa	attagccagg	cgtggtgtcg	34740
cgcacctgta	attgcagcta	cttgggaggc	tgaggcagga	gaattgcttg	gaccaggag	34800
gcggagggtg	tggtaagcca	agatcgcgcc	attgcactcc	agcctgggca	acaagagtga	34860
aactctgtct	cagaaagaaa	gaaaagcaaa	acagtatttc	atttaattaa	aaatgccttg	34920
tattatacca	gtaagagtaa	aaaacaattt	aattaaagta	tgatatatgc	tcttttcatc	34980

acttagaatt	tgaatttttaa	tcttattgaa	gaggctgttt	tagttttatt	tagatatggt	35040
tataacaaca	atattaatat	ctgtgaccaa	attggtgcat	gcatgggcaa	tgaaaagcta	35100
tgttgtgact	gcagcatggt	ggatagcatt	agtaagttag	cattggttgt	aaaatgaatc	35160
ccaatttcaa	agattttgaa	acgtggggaa	aaaatatatt	agaccccatg	aaataagacc	35220
tgaaatctat	gaaataccac	agtaggttgg	aaatcatcat	gagaaactgt	aactattttt	35280
ttttataggt	attatatgta	ttctttgaga	cagggtctca	ccctgttgcc	caggctggaa	35340
tgctgtgatg	tggtcacctc	tcactgcaac	atccacttcc	tggggctgaa	gtgatccttg	35400
caactcaccc	tcccatgctc	ggctaatttt	tttttttttt	tttttaagta	gagaccagat	35460
ctctttatgt	tgcccaggct	ggtcttgaac	tcctgggctc	aagtgattct	gctttgtctg	35520
cccagagtgc	tgggattaca	gaagtgagcc	acaacacctg	gcctaaagat	agtattttta	35580
aaatttcttc	tgttttcttt	cttttgggta	tacattttct	tctttttttt	aaaaaaaaat	35640
ttctgtataa	agactgtgac	tcccatgaa	agtagtttgg	gtgataattt	atcgtgaaac	35700
aatattagtt	gggaaatgta	gatattttta	ttaatttttt	tctttcaggt	gaacagtgtg	35760
aaagagctag	acgccttcaa	gctaaaatga	tgactaattt	ggttatggcc	aaggaccgct	35820
tacaacttct	aggtatcaat	taatgtataa	tttgatgtgg	gatgtattgg	aaatgtgtgt	35880
tcaatgaaac	tttaatttgt	agaaagaaat	agatcagtga	ttgaaaatgt	ggtccaggct	35940
ttttaacata	aagaaaacgt	ataacatata	caaaaacaga	agagcataat	ggacttcttt	36000
ttttttttga	gatggacgct	tgctctgtca	cccaggctgg	agtgcagtgg	cgtgatcttg	36060
gctcactgca	agctctacct	cccaggttca	cgccattctc	ctgcctcagc	ctcctgagta	36120
gctgggacta	caggcacccg	ccaccacgcc	ccgctaattt	ttttgtagtt	ttagtagaga	36180
cagggtttca	ccatgttagc	cagaatggtg	tccatctcct	gacctcgtga	tccaccgcc	36240
tcggcctccc	gaagtgctgg	agttacaggc	gtgagccaac	gcgcccagcc	gagcataatg	36300
aacttctaatt	tattcttctc	tcagctttga	taattatcac	ctcactatca	gtcttgattt	36360
atttatacca	agggctagca	aactgcagct	gatgggccaa	cccagctcac	tgcctattac	36420
tataaataag	gtttttattgg	cacacagcca	ttcctttcct	ttaggtattg	tctatggctt	36480
tcatgctacg	gtgacagaat	tgaatagttg	atacagagac	aacactgccc	tcaaaaccta	36540
acatattttac	tgtctggccc	tgagaaagcc	tgccatcttc	acttgacatc	ccacttcact	36600
gtcccccaac	ttcttattat	tttgaaacaa	attattgact	tattttccat	ttataaatat	36660
ttcatattat	acctctaaaa	gataaggatg	tcaaaataaa	atcacagaaa	taccattatc	36720
acacctaaaa	aataactcca	aatttctttt	aacattatcc	aaaaaaactg	acccccccaa	36780

aatgcaattt	ccaaatttct	aattttctca	aaatttaata	atttttacta	ttttttaata	36840
at ttgtttgt	atcaggatct	taaaaagatc	cgtatgtctc	ttttaatgta	caggttcatc	36900
tg tttttcat	tttcttccaa	tttatttggt	gatgaaacca	ggctcttgtc	atgtaatatt	36960
tcctacagtc	tgacttttgc	tgcatgcatc	ttttgtggta	tg ttttaaca	tg tttctctt	37020
tcctcccacg	ttaaatgata	gttggatata	gagccttgat	cacattgaaa	gttgattttt	37080
at tttttaat	gattagacta	cttcctaggt	ggttgtgttt	gtgtgttcat	cattagttgt	37140
tccatcttgt	gagggttagca	gcagttaact	atgactgatg	tccagtgata	ctttaattct	37200
gtcattcttt	cagcattcat	caactgtaat	agaaactttt	acttgtcttc	tg tttgatgg	37260
cgtagggttt	aaatatatat	taataagtta	aatctaccac	tcaaaaaaag	tagaacaaaa	37320
cctactacat	atgagtggct	ctaccatacg	ttgggtggatg	gaaatagtag	atatttggtta	37380
agcaggtaaa	ccttatttat	ttgtagccct	gtgacttggg	gtaagttatt	ttgcatgtct	37440
gggtcttttt	tatctaattt	gtaatatgaa	gctattatta	ttattattat	tattattatt	37500
attattatta	ttattattat	tattcgagat	ggagtcttgc	tttgttgccc	aggctggagt	37560
gcagtgggtc	aatctcggct	caatgcaccc	tccacctcct	gggttcaaac	tattctcgtg	37620
attcagtctc	ccgagtagct	aggattacag	gggcctacca	ccacacctaa	tttttgtatt	37680
tttagtaggg	acgggggttc	accatgttgg	ccaggctggt	ctcaaactcc	tgacatcaag	37740
tgatccatcc	acctcagcct	cccaaagtgc	tgggataggc	cgggcacggt	ggctaacgcc	37800
tgtaatccca	gcacttttgg	aggccgaggc	gggtggatca	cgaggtcagg	agattgagac	37860
catcctggct	aacacggtga	aaccccgctc	ctagtaaaaa	aatacaaaaa	actagccagg	37920
tgtggtggcg	ggcacctgta	gtcccagcta	cttgggaggc	tgaggcagga	gaatggcatg	37980
aacccaggag	gcggagcttg	cagtgagccg	agatcatgcc	actgcactcc	aggctgggca	38040
aaagagcgag	actctgtctc	aaaaaaataa	aataaaataa	atagataaat	aataataata	38100
aagtgctggg	attacaggca	tgagccacca	cgcccagctg	aagctaatat	tattagctaa	38160
taataatatc	tcatattttt	gaccattgaa	ttattaaagg	aacctaaacc	atagtagtaa	38220
gtgctttaaa	aatgttggtt	tttattttatc	tg tttacctg	actgtcatca	cctcctttgc	38280
ttaggtagaa	agatatattc	ataaacacat	acatatacat	tttattcatt	ctttaattca	38340
tttactccta	tatggtcatt	gtgtttgtgg	gatttttgcc	cgcacatagt	taaaaatgca	38400
gccgttactg	ctttcagatg	agtaacaagg	tagtgttccc	tggtctctgt	ggctgacagg	38460
at ttgccttc	ttcctttcta	aatggaggtt	attactgtgt	cagatataat	taaatagtgt	38520
attaaagttt	gttgcaataa	aataattgat	ggttctaatt	ggtactttct	acgtgtttta	38580

tcttttaaagc	ttttcagtg	atatataaag	tatatatcat	acagaataaa	tttgattgtg	38640
gagcattttg	taacatcttt	caaaattaat	aagagttggt	ttttatgttt	tgtttgatgt	38700
tatgatatag	gacattagaa	gtattaatat	caattagaga	ctcatctttg	aatgtgactt	38760
tgtactttct	tatttggtgt	agtagaggag	agaacaaaaa	gaagatatgt	aatgtaatat	38820
gaaactaggc	atttaaagat	ttaacgtttt	ggatatttta	aagttggtgt	ctgttttcac	38880
cctcaaaaat	gctattacct	atttatgaaa	tatctttaaa	agtgtggagt	ggtagatggg	38940
aaaggtgaca	tcggctgggc	gcggtggctc	acacctataa	tcccagcact	ttgggaggcc	39000
aaggtgggcg	gatcacctga	ggtcaggagt	tcacgaccag	tctggttaaca	tggtgaaacc	39060
taatctctac	taaaaataca	aaaaattagc	caggctgggt	gtacgcacct	gtaatccag	39120
ctactcagga	ggctgagtca	ggaggatggc	ttgaaccag	aaaatggagg	ttgcattgag	39180
ccaagatggc	accactgcac	tccagactgg	gtgacagaga	gagactccgt	ctcaaaaaaa	39240
aaaaaacgcc	gggcactgtg	gctcacacct	gtaatcctag	cactttggga	gactgaggcg	39300
agcagatcac	ttgaggtcag	gagttcgaga	ccagctgggt	gcctgggcaa	catagtgaaa	39360
ccccgtctct	actaaaaata	caaaaaaatt	agccgggcgt	agtggcgggc	gcctgtagtc	39420
ccagctactt	gggaggctga	ggaaggagaa	tggcgtgaac	ccgggaggcg	gagcttgtag	39480
tgagccgaga	tcccgccact	gcactccagc	ctgggtgaca	gagcgagact	ccgtctcaaa	39540
aaaaaaaaaa	aaaaaaaaaa	tagccgggca	tgggtggcggg	cacctcta	ctcacctact	39600
tgggaggttg	aggcatgaga	attgctcgag	cctgggaggc	agaggttgca	gtaagcagag	39660
accacaccac	tgactcttag	cctggatgac	agagttagac	tttgtctcaa	aacaaaaaca	39720
aacaaaccaa	accacctctt	gtccactcct	aatctccctt	actgtgtttc	acttattttt	39780
ttcccatata	ttgtcacct	tctaataaac	tgtgtaactt	attttttatg	tttattgtgt	39840
actatctgtc	ttctccctgt	tggaatataa	actctatgga	ggagggatcc	ttgtctatat	39900
cccaagtgcc	tacaacagtg	tctagcacat	gctagttacc	gaataaatat	ttgccaagtt	39960
aatgaagttg	tttataaccc	ttggcactat	acctgataca	gtgtagggac	ttaataaatg	40020
aatgttaa	gaagctatct	taaaaatata	tatatatata	taacagccag	atacgccaat	40080
ttgccgacat	caaaattaat	cttttcatct	attgcaggtt	ttacagttaa	gctttgtctc	40140
aaggttcact	atcactaaac	caaaaataaa	taatatggat	ggtcttttcc	catttttaac	40200
agagaggcta	catcctctca	aaccagtcag	tattaggaat	tcccaggatc	attctctcca	40260
ttttccatcc	taattagata	cctatctggc	accacatctc	tgaaaagctc	caatttttaa	40320
ccagcacata	ataaacaat	atgttactat	taatacacgt	ttgtacacaa	gtcattttta	40380

ttgtcttagg	tatatccact	cccagacctt	ctgagttatt	gaaagacagg	aatttgtgat	40440
gtcacctctt	ttttcttttt	tttgagacgg	agtcttgccct	tatcacccag	gccagagtgc	40500
aatggcgtga	tctcgggtca	ctgcaacctc	tgcctcccgg	gttcaagcga	ttatcctgcc	40560
tcatcctcct	gagtagctgg	gattacaggt	ggccctagtt	tgggtttttg	gttttaaaca	40620
ggatgggtctc	ctctttgaga	agttgccatg	agttagccat	gatgatacct	gggacaataa	40680
tgggtccacag	agaagtaaga	gccaatacaa	cgccttaaga	gaaagtacct	ggtatgtttg	40740
tgaagcagca	aagaggccaa	tgtgttccag	caaggagatt	atagtagaga	tgaagtcaga	40800
gaactcatgc	ggagagggat	tgtcacaaca	cgcagggctg	taaaggccac	catgaggact	40860
tgagctttta	ccgcctgtga	aatggggagc	agttacaaag	atttttaaga	aagcagccca	40920
gtgcagtggc	tcatgcctat	aatcccagct	ctttgggagg	ctgaggtggg	tggatcacct	40980
gaggtcagga	gttcgagacc	agcctggcca	acatggcgaa	accttgtctt	tactaaaaat	41040
agaaaaatta	gctgcggtg	gtggccatgc	ttgtaatccc	aaatactcgg	gaggctgagg	41100
cagaagaatc	acttgaaccc	gggaggcgga	ggttgcaagt	agccgaggtt	gcgccactgc	41160
cctccagcct	gggcaacaaa	gcgagactcc	gtctcaaagc	cagaaaaaaa	gatttttagg	41220
aaaggaatga	aatactctga	cttaaatttt	agaaggatca	ctctggttgc	tatgttgcat	41280
tatagactct	aagggtagaa	ccagggagag	caattaggac	actgttacag	taatctagga	41340
gaaaagtgat	agaggcttgc	accagggtag	tatcaggaga	aatggtaaca	agtggttgga	41400
ttatgattat	agttgaaggt	cgaacaacta	gatttgctga	tagattggat	aaaatggtcc	41460
tctctttatg	gatctcataa	tacagtagga	gaaacagaca	cataaatagt	catcactgca	41520
ggatgggtat	tagaaatatt	ttgcaaggg	gtaaaaggtg	ctaagaaagc	tttcttgtgt	41580
gacaaagctg	gttaaagaaa	gaaaagaaag	cttcctagaa	gtgaaaccta	aaatatacac	41640
ttagataaat	gaataaaagt	tatgaaagaa	tgtgagggct	gggcacagt	gctcacacct	41700
gtaatcctag	cactttggga	ggttgagtcg	ggggtattgc	ttgagtccag	gagtttgaaa	41760
ccagcctgga	cagcgtactg	agaccccatg	tctaaattta	aaaaaaaaaa	aaatagccag	41820
gagtgggtgat	gtgcacctgt	agtcctagct	acttacttgg	gaggctgagg	caggaggatt	41880
gcttgagctc	aggaggtcga	gactgcagtg	agccatgagt	acaccactgc	actctagtct	41940
gggtgacaga	gtgagaccct	gtctcaaaga	aaatgaaaag	aaaaagaatt	tgagaaaaga	42000
tatgtcagga	aaatgtgatg	acatgaacaa	atactcaaag	gcaagaaaag	catggtgagt	42060
aggggagggg	acaggttgag	acatgaagtt	gggattgccg	aggcgtaaga	ggtataggca	42120
gatcatggat	ggcctgggtt	gtcataggaa	ggagcttggg	ctcttatctg	tgggcaatgg	42180

gaagccacta	aagggtttta	agtagaagag	tgttatatgg	taagggttttc	ctctccagta	42240
gatcactcag	atgactgtgg	gggttg gatg	tgaggaggta	aagcagcaaa	gttactgctt	42300
taattttaa	cattgattaa	tctgtacaat	cctagtcatt	ccaaaaagaa	acattagtc	42360
tttaacagaa	gtgaattggg	tctggg cgcg	gtggctcagc	cctgtaatcc	cagtactttg	42420
ggaggccgag	gcgggcagat	cacctgaggt	caggagtttg	agacaagtct	ggccaacatg	42480
gcaaaacctt	gtatctacta	aaattacaaa	aattagctgg	gcatgggggc	acccacctgt	42540
aatcccagct	attcgggagg	ctgaggcagg	agactctctt	gaaccataa	ggcagagggt	42600
gcaatgagct	gaggtcgcac	cactgcgcctc	cagcctgggc	gacagagtga	gactccgtct	42660
ccaaaaaaa	gaaaaaatat	tggagcagtt	tcacagatgc	tgtttactgt	tatactgtat	42720
gtgtctatga	ctcctcctcc	aagaaaaaaa	aatgaattg	gagcagggttc	acagaagcaa	42780
tgtactgtta	cactgtatgt	gtctgtttct	acacatataa	atctgaattc	tgtgtacacg	42840
aaaagaaatc	ctatagcttt	ttattcctag	ctataaaaac	taagaatata	atttctttct	42900
ttttattaca	taatgaatgg	ttctgttaac	tttttg ttaa	aatttcattg	aggggagtat	42960
taattcacat	ttatgtacgg	atgtctacat	ttacaaatca	gtgtattttt	tgatttatgc	43020
ttttactgag	acaaaggggt	tctgtctcag	catggtcatt	taaagagttt	atcattgaga	43080
aaaatcagat	gaccaacctg	ttagctcaaa	aaaaaaaaaa	cctccaaggt	atattgtatc	43140
agccagttct	aggatacaaa	agccatgcag	tactttgtgt	tttgtgccaa	aaagggtagc	43200
tgctatttga	cctgtcccaa	aggcatgtgt	ggttgtagccg	taaaccaagc	atggtacctg	43260
tttgtcaaac	tttagaaatg	aaagtttaag	agagttaata	tatagggtgct	gcatttttta	43320
tgtattcatt	gacttgctgg	tacagaagaa	aagaatcaat	tatgattcag	cacaatactc	43380
cacttgggga	agagagtgca	gcagtagttt	agagtgtcag	ggatcaaact	gctaccttct	43440
tgggcttcag	ttgctagact	taagagaccc	agatcttggg	aggggttttt	gttgttattt	43500
gatgtggggg	taaggttgaa	atactccctt	ttaattgatt	aatataaaaa	ataggtaatg	43560
tagacgtatg	tcacaaaatt	ttaacaagtt	tgagtgtgaa	aagcatccat	cccattctat	43620
gccctggcta	cccatctccc	ctccctagag	gcggctgtta	ttatcagttt	ctctgggtcc	43680
tttttagagg	taatcttttt	gtatgcaact	gtgcacatgt	atgtttacac	aaatggtagg	43740
gtattgtata	tatatacagt	tactgcagta	tgcttatttg	actttgcttt	attgctttta	43800
ttttctgcct	gacatatgtt	aatgtggcca	aatagggcat	ttcccagaca	atccttatag	43860
tttcattctg	tgctacttta	aatgctgttc	tccttccctg	aaatgtcctt	tttttttttc	43920
ttttcttttt	ttttttttga	gacagagtct	cgctctgtca	cccaggctgg	agtgcagtgg	43980

cacaatctca	gctcactgca	acctccacct	ccccggttta	agccattccc	ctgcctcagc	44040
ttcccagagta	gttgggacta	caggcgcggtg	ccaccacgcc	cggatacttt	tttgtttggt	44100
tgtttttagta	gagacagggt	ttcactgtgt	tagccaggat	cgcctcaatc	tcctgacctc	44160
ctgatctgcc	tgcctcagcc	tcccaaagtg	ctgggattac	agacgtgagc	cactgcgctt	44220
ggctgtcctt	tcttttaatc	ttttcaaatt	ctagccagtt	catagcccac	ttgctctgta	44280
aagctttttc	agtgatcttt	cttttgtacc	atgcagtctt	ttaaaaaatc	tacagtttta	44340
ttataagttg	ctaattatat	ttaataaata	cctaaatggt	agaaattgat	tcatttttta	44400
atctccgtag	tgccaggaac	tgtggctcat	gtctgtaatc	cctgcacttt	gggaggctaa	44460
ggcgggagga	tcgcttaagg	ccaggagctt	gagcccagcc	tgtggaacat	attgagaccc	44520
cgtctctaca	aaaaataaat	tagctgggca	tagtgggtgc	tgtatgtagt	gtcaactact	44580
caggagactg	aggcagaagg	ataacttgag	ctcaggaggt	tgaggccaca	gtgagccatg	44640
atggtgccac	tgtacttcag	cctgggtgac	agagcaagac	caagaccgat	aattatctcc	44700
atggtatttg	ttttggacgt	attagatatt	cagtaaatat	tttctggtga	tgatagtgat	44760
atctaactac	ctcaggttac	ctaacctgaa	taaggcctat	ttttaacata	gccattcact	44820
tgtaaatctc	tgttattttc	gtgacttcat	ttttactaat	tatatatttaa	aagataatct	44880
tatataaatg	caaaaaacttt	ttatcatgta	acaatctggt	aacaccttga	gtaatttgct	44940
atttcacatg	cacattttat	ttgttcatta	tcttttttct	ttttgtttat	tttttctggt	45000
ttttaccttc	tctgttgcat	agagaagatg	caaccagttt	tgccattttc	caagtcacaa	45060
acggacgtct	ataatgacag	tactaacttg	gcatgccgca	atggacatct	ccagtcagggt	45120
gggttttaggt	taactaacat	aaaataataa	agcttgcatg	caaagtaaga	gtcttactta	45180
acctgataat	gttgatttga	ttttataatg	gtaggtttta	ttgttcattg	tttcacaggg	45240
ctgtgttgaa	aatatagtag	ctttatcatc	atagtatata	aacatgcaac	aagtcaggta	45300
tagtcattct	aaacttgctt	tatttgcata	ttttgcaacc	tctgggtttt	cagattataa	45360
aatatgaagg	tatacatctc	cgtaagtgat	taaaataatc	tttttttctg	ttgtggcttt	45420
tgtgttttta	tatggatagc	tatactgatt	tttttcatga	aatagtgttt	tctaagacac	45480
attttaatct	tttatgtatc	aatttttatt	gcatatcaat	tttcattttat	agtatatgtg	45540
aatgtcttat	agtacttgat	agaagaaatt	tgaaacttca	catagtgagg	agaaacatta	45600
cagtattatt	tgctatggat	taactctatc	cttcataaac	ttttggatat	caacttctac	45660
agaagattaa	atgtctttct	gttttggttt	ctttgggtag	gggactctcc	atttcttagg	45720
ttctctgaca	tctgagctag	ctttttgttt	aattaaaact	ttttttttag	ttgacagata	45780

atgtacctgt	tcatggggta	cacagtgatg	ttttgataca	tatagtatat	ggtgattaga	45840
tcaaggtaat	tatcatttcc	atcatcccaa	acatttatca	ttccttgtgt	tggaaacatt	45900
cagtattttc	cttctagcat	ttgaaatgat	ataatatatt	atattgttac	ctataatcat	45960
cctgcagtga	tgtagaacaa	tggagcttac	tctttctatc	tagctatacc	tttgtatcct	46020
ttccctaacc	ttttcagcct	ctagtatcct	ctgttctact	ttttatttct	atgagattaa	46080
ctttttaaaa	ataatattct	gtgtgtatat	ataccacatt	ttctttattc	attcgtctgt	46140
tcttggacac	ctaggttgat	tctgtatcct	ggctgttggt	aatagtgtgt	cagtaaacad	46200
ggggtgtaga	tgtctgttcg	atatgatgat	tttctttcct	ttggataaat	tcccaatagt	46260
gggattgggt	ggtcatatgg	tagttctggt	tgtagttttt	gaggaacctc	catactcttc	46320
tccatagtag	ctgtactagt	ttagattccc	aaaagtagcg	tataagagtt	cactttctac	46380
acatcctcac	cagcatttgg	tacttttttt	gtctttttga	taatagccat	cctaactggg	46440
gtaagatacc	ttgttgtggt	tttgatttgc	atttccctca	taattaacga	tgttgagcat	46500
tttttcatat	ctttgttttt	gacggtccag	ctagttttat	tggttacttt	tttttttttt	46560
ttttagcaat	agtttcatgt	gcagggctca	aattatatatt	aatattactt	ttttcaaatt	46620
ataccagaac	aaccttaggt	taggctataa	aacaactgcc	ttctttttct	tttggcataa	46680
taattatata	tatattttta	attacacttt	acgttctagg	gtgcatgtgc	acaacatgca	46740
ggtttgttac	atatgtatac	aagtgccatg	ttggtgtgcc	gcacccatta	acttgtcatt	46800
tacattaggt	atatctccta	atgctacccc	tccctctctc	cctcaccoca	cgacaggccc	46860
cattgtgtga	tgttccccac	cctgtgtcca	agtgttctca	ttgttcagtt	cccatctgtg	46920
agtgagaaca	tacggtgttt	ggttttctgt	ccttgcgata	gtttgctcag	aatgatgggt	46980
tccagcttca	tccatgtccc	tacaaaggac	atgaactcat	cctttttgat	ggctgcttag	47040
tattccatgg	tgtatatgtg	ccatattttc	ttagtccagt	ctatcattga	tggacatttg	47100
cattgattcc	aagtctttgc	tattgtgaat	agtgccacag	taaacataca	tgtgcatgtg	47160
tctttatagc	agcatgattt	ataatccttt	gggtatatac	ccagtaatgg	gatggctggg	47220
tcaaagtgta	tttctagttc	tagatccttg	aggaatcgcc	acactgtcct	ccacaatggt	47280
tgaactagtt	tacgttccca	tcaacagtgt	aaaagtgttc	ctgtttctcc	acatcctctc	47340
cagcacctgt	tgtttcctga	gtttttaatg	atcgccattc	taactgggtg	gagacgggat	47400
ctcattgtgg	ttttgatttg	catttctctg	atggccagtg	atgatgagca	ttttttcatg	47460
tgtcttttgg	ctgcataaat	gtcttctttt	gagaagcatc	tgttcatata	ctttgcccac	47520
tttttgatgg	ggttgtttga	ttttttcttt	gtaaattttt	ttaagttctt	tgtagattct	47580



1

2



ggatatcagc	cctttgtcag	atgggtagat	tgcaaaaatt	ttctcccatt	ctataggttg	47640
cctgttcact	ctgatggtag	tttcttttgc	tgtgcagaaa	ctcttttagtt	taattagatc	47700
ccatttgtca	attttggctt	ttgttgccat	tgcttttggg	gttttaggca	tgaagtcctt	47760
gcccattgcca	gtgcccattgc	cagtgtcctg	aatgggtattg	cctagattttt	cttctagggt	47820
ttttatagtt	ttagaactaa	catttaagtc	tttaatccat	cttgaattaa	tttttgtata	47880
aggtgtaagg	aagggatcca	gtttcagctt	tgtacgtttg	gctagccagt	tttcccagca	47940
ccatttatta	aataggaaat	cctttcccca	tttattgttt	ttgtcagggt	tgtcaaagat	48000
cacatgggtg	tagatgtgtg	gtattatttc	tgagggtctt	gttctgttcc	attggtctat	48060
atctctgttt	tggtatcagt	accatgctgt	tttgattact	gtaccttcgt	agtatagttt	48120
gaagtcagggt	agcacgatgc	ctccagcttt	gttcttttgg	cttaggattg	tcttggaat	48180
gcaggctctt	ttttggttcc	atatgaactt	taaagtagtt	ttttccaatt	ctgtgaagaa	48240
agtcatttgt	agcttgatgg	ggatggcatt	gaatctataa	attaccttgg	gcagtgtggc	48300
cattttcatg	atattgattc	ttcctatcca	taagcatgga	atgttcttcc	atttgtctgt	48360
gtcctctttt	atttcattga	gcagtgggtt	gtagttctcc	ttgaagaggt	ccttcacatc	48420
ccttgtaagt	tggattccta	ggatattttat	tctctttgaa	gcaattgtga	atgggagttc	48480
actcatgatt	tggtgtttt	tctgttattg	gtgtatagga	atgcttgtga	tttttgcaca	48540
ttgattttgt	atcctgagac	tttgctgaag	ttgcttatca	gcttaaggag	attttgggct	48600
gagatgatgg	ggttttctaa	atatacaatc	atgtcatctc	caaacaggga	caatttgact	48660
ttcttttcct	aattggatac	cctttatttc	tttctcctgc	ctgattgccc	tggccagaac	48720
ttccaacact	atgttgaaca	ggagtgggtga	gagagggcat	ccctgtcttg	tgctagcttt	48780
caaaggggaat	gcttccagtt	tttgcccatt	cagtatgata	ttggctgtgg	gtttgtcata	48840
aatagctctt	attattttga	gatacatccc	atcaatacct	agttttattga	gagtttttag	48900
catgaagggc	tgttgaattt	tgtcaaaggc	cttttctgca	tctattgaga	taatcatgtg	48960
gtttttgtct	ttggttctgt	ttatatgatg	gattacgttt	attgatttgc	atatgttgaa	49020
ccagccttgc	atcccaggga	tgaagccaac	ttgattgcgt	tggataagct	ttttgatgtg	49080
ctgctggatt	cggtttgcca	gtattttatt	gaggattttt	gcgttgatgt	tcatcaggga	49140
tattggtcta	aaattctctt	tgttgtgtct	ctgccaggct	ttggtatcag	gatgatgctg	49200
gcctcataaa	atgagttagg	gaggattccc	tctttttcta	ttgattggaa	tcatttcaga	49260
aggaatggta	ccagctcctc	tttgtacctc	tggtagaatt	cagctgtgaa	tccatctggg	49320
cctggacttt	ttttgggttag	taggctatta	attattgcct	caatttcaga	gcctgttatt	49380

ggctctattca	gggattcagc	ttcttcctgg	tttagccttg	ggaggctgta	tgtgtccagg	49440
gatttatcca	tttcttctag	atcttctagt	ttatttgagt	agagggtgtt	atagtattct	49500
ctgatggtag	tttgcaattc	tgtggaatcg	gtggtgatat	cccccttata	atctcttatt	49560
gcatctatct	gattcttctc	tcttttcttc	tttattagtc	ttggtagcag	tctatcagtt	49620
ttgttgatct	tttcaaaaaa	ccagctcctg	gattcattga	ttttttgaaa	ggttttttgt	49680
gtctctatct	ccttcatttc	tgctctgata	ttagtatttt	cctgccttct	gctagctttt	49740
gaagggtgtt	gctcttgctt	ctctagttct	tttaatgggt	atgttagggg	gtcaatttta	49800
gatctttcct	gttttctctt	gtgggcattt	agtgtataaa	atttccctct	acacactact	49860
ttaaatgtgt	cccaaagatt	ctgatatgtt	gtgtctttgt	tctcgttggt	ttcaaagaac	49920
atctttatct	ctgccttcat	tttcttatat	acctcagtagt	cattcaggag	cagggtgttc	49980
agtttccatg	tagttgtgtg	gttttgctgt	agtttcttaa	tcctgagttc	tagtttgatt	50040
gcactgtggc	ctgagagaca	gtttgttgta	atctctgttc	ttttacattt	gctgaggagt	50100
gcttttagttc	caactatgtg	gtcaattttg	gaatagggtg	gggtgtgggtg	tgagaagaat	50160
gtatattctg	ttgatttggg	gttttagagtt	ctgtagatgt	ctattaggtc	cacttggtgc	50220
agagctgagt	tcagttcctg	gatctgtctt	gttgatctgt	ctaataattga	cagtgggggtg	50280
ttgaagtctc	ccagtattat	tgtgtggggag	tctaagtctc	ttttagagtc	tctagggact	50340
tgctttatga	atctgggtgc	tcctgtattg	gggtcatata	tatttaggat	agttagctct	50400
tcttgttgaa	ttgatccctt	tagcattata	tgatggcctt	ctttgtctct	tttgatcttt	50460
gttggtttta	agtctgtttt	atcagagagt	tggattgcaa	acctgctttt	ttttgttttc	50520
catttgcttg	gtagatcttc	ctccatccct	ttattttgag	cctatgtgtg	tctctgcacg	50580
tgagatgggt	ttcctgaata	cagcacactg	atgggtcttg	actcgttata	caatttgcca	50640
gtctgtgtct	tttaattgga	gcatttagcc	catttccatt	taagggttaat	attgttatgt	50700
ttgaatttga	tcctgtcatt	atgatgttag	ctggttattt	tgctcgttag	ttgatgcagt	50760
ttcttcctag	cctcgatggg	ctttacaatt	tggcatgttt	ttgcagtggc	ttgtaccggg	50820
tgttcatttc	catgttcagt	gcttccttca	ggagctcctg	taagcaggcc	tggtagttac	50880
aaaatctgtc	agcatttgct	tgtctttaaa	ggattttatt	tctccttcac	ttatgaagct	50940
tagtttggtc	ggatatgaaa	ctctgggttg	aaaattcttt	cctttaagaa	tgttgaatat	51000
tggcccccac	tctcttctgg	tttttagagt	ttctgccaag	agatcagctg	ttagtctgat	51060
gggcttccct	ttgtgggtaa	cccgcctttt	ctctctgggt	gcccttaaca	ttttttccta	51120
catttcaact	ttggtgaatc	tgacaattat	gtgtcttgga	gttgctcttc	ttgagtagta	51180

tctttgtggc	attctctgta	tttctgaat	ttgaatgttg	gcctgccttg	ctaggttggg	51240
gaacttctcc	tggataatat	cctgcagagt	gttttccaac	ttggttccat	tctccccgtc	51300
actttcaggt	acacccgtca	gacatagatt	tggtcttttc	acatagtcct	atatttcttg	51360
gaggttctgt	tcgtttcctt	ttactctttt	ttctctaaac	ttctcttctg	gcttcatttc	51420
attcatttga	tcttcaatca	ctgataccct	ttcttccact	tgatcgaatc	ggctactgaa	51480
gctcatgcat	gcatcacgta	gttttcgtgc	catggttttc	agctccatca	ggcca'tttaa	51540
ggtcttctcc	atgctgttta	ttctagttag	ccatttgtct	aatctttttt	caaggttttt	51600
agcttctttg	caatggtttc	gaacatcctc	ctttagctcg	gagaactttg	ttattaccca	51660
tcgtctgagg	cctacttctg	tcagcttgtc	aaagtctttc	tctgtctagc	tttgttccgt	51720
tgctggtgag	gagctgtggt	cctttggagg	agaagaggcg	ctctgaattt	tagaattttc	51780
agcttttctg	ctctggtttc	tccccatctt	tgtggcttta	tctacctttg	gtctttgatg	51840
atggtgacgt	acagatgggg	ttttggtgtg	gatgtccttt	ctgtttgtta	attttccttc	51900
taacactcag	gaccctcagc	tgcaggtctg	ttggagtttg	ctggaggtcc	actccagacc	51960
ctgtttgcct	gggtatcacc	agcggaggct	gcagaacagc	aaatgttgca	gaacgacaaa	52020
tgttgctgtc	tgatccttcc	tctggaacct	tcgtctgaca	ggggtaccca	ggtatatgag	52080
gcgtcagtca	gcccgtatgg	ggaggtgtct	cccagttagg	ctacttgggg	atcagggacc	52140
cacttgagga	ggcagtctgt	ccgttcgccg	atctcaaact	ccatgctggg	agaactacta	52200
ctctctttag	agctgtcaga	cagggacctt	taagtctgta	gaagttactg	ctgccttttg	52260
ttcagctatc	ccatgtcccc	agaggtggag	tctacagagg	cagtcaggcc	tccttgagct	52320
gtggtggact	ccaccagtt	caagcttcct	agctgctttg	tttaccact	caagcctcag	52380
caatggcaga	tgccccgcc	tccagcctct	ttgccgcctt	gcagttcgat	ctcagactgc	52440
tgtgctagca	gtgagcgagg	ttccgttggc	atgggaccct	ctgagccagg	catgggatat	52500
aatctcctgg	tgtgccgttt	gctaagacca	ttggaaaagc	ccagtattag	ggtgggagtc	52560
tccctatttt	ccaggtacca	tatgtcacgg	cttcccttgg	ctaggagagg	gaattcccca	52620
acccttgcg	cttcccaggt	gaggcaatac	cccgccttc	agctcacact	atgtgggctg	52680
caccactgt	ctgacaatcc	ccagtgaat	gaaccagtt	cctcagttgg	aaatgcagaa	52740
atcagctgtc	ttctgtgtcg	ctcacactgg	gagctgcaga	ctggagctgt	tcctatttgg	52800
ccatcttggg	accctgcctt	cttcattcat	atgtaataca	aaacttctaa	ggttttagtg	52860
gagaagagat	agagtaaaag	gatttctata	gacagaagaa	acagttggtc	atcaactctt	52920
tcccttgtgg	cttcacatct	cccctaagga	cttcttatgt	tggtttggtc	ttacagtata	52980

gtcagtgga	gtttccctct	tccattcct	tcctctgtct	gatttaaaat	gctgtttcaa	53040
gtatcgatat	aagtattttg	cctgtttctt	ttagcgtggc	tgtgaagggc	tgacattttc	53100
agaaggcact	tactgaaaaa	aaaaaaacaa	agaaatgtaa	gagtcacatca	catataaata	53160
gttaagtttc	taaaatatgt	atttgagatc	ccagtaattc	tactaggata	aatagcaaaa	53220
attctccagc	cctgaagagt	tggtctgtct	ttcctttcct	tggtatcttg	attctctttt	53280
atttcatttc	acatgctaga	ccccatcatg	ttttcctgct	gcattcccca	ctccacccctc	53340
caaagcaatg	ttctcttcct	gcaacctgtt	tgaaaaaatt	gaactttgtc	ttattttcata	53400
tccttagtac	aagttgaata	tcctaataat	gaaatctgaa	atccaaaatg	ctctaaaatt	53460
cgactctttt	ttttttgaga	tgaagtcttc	ctctgtcact	caggctggag	tgcggtggcg	53520
caatcttggc	tcactgcaac	ctccgcagcc	tcccctaacc	ccatccctgt	gggttcaagc	53580
gattctccca	cttcacctc	ccaagtagct	gggactacag	gcacctgcca	ccacacccaa	53640
ctaatttttt	gtatttttag	tagagatgag	gtttcaccat	gttgggtcaag	ctagtctgag	53700
actcctgacc	ttaagtgatc	cacctgccct	ggcctcccaa	agtgttggga	ttacagtcac	53760
gagccactgt	gcctagccaa	atatccaaaa	ctttttgagc	gctgacatga	tgctcaaagg	53820
aaatactgga	gtattttgca	ttttggattt	ttgggttaag	gatgctgaac	cagtaagtat	53880
aatgtaaaca	ttgcaaaatg	caaaaaaatg	taaaaaccct	aagcagttct	ggtcccaagc	53940
atttcgcata	agggatactc	aaccataat	cttttctttt	ctgttttctg	gttgggaagg	54000
catattggct	ttatagctaa	tttacacagc	attgatgttt	aatacacagt	gagtccaagt	54060
agatcactca	gacctattag	tagtttatta	gtgtcactca	cttctggaac	attctgtgat	54120
gttttacttg	gatgagttct	ttcacctctc	ttgggaaata	gtcataccaa	agtctgctta	54180
ttactacaat	gtgttatcta	ttatacattg	tccaattttc	tactcaaaat	tactagacag	54240
gcaaagaaaa	agtaaagtat	aacccttact	caagaaaaaa	aagcaatcag	tagaaactgt	54300
gagtgggccc	agatattgga	tttagcagac	aaagacttca	aagcagctat	tataatatgt	54360
ttaaagaatt	gaaagaaaat	atggtatcag	ttaaacagga	aatctaagta	gatgatataa	54420
actagacaat	aaaaataata	ttctaagtt	gaaaagtgtg	gttactgaaa	ttaaaaattt	54480
acagacaaca	gcctcaacag	cagattagag	atagcaaaag	aaagattcag	tgactttgaa	54540
tgaggtctg	tagaaattac	taaattgatg	actctcacgt	agcaactttc	accgtagtt	54600
tggtttcata	tacaatgctt	taactctttt	tgctattttt	tctcttactt	tctgtgatgg	54660
aaacattttg	tgctttgatt	ttaatgggtg	tatcacagtt	atatacatct	ttacaaactc	54720
atcaaattgt	atactttatt	tatatattta	tttttttttg	agatagggtc	tgactctttc	54780

tcccaggctg	gagtacagct	gaaattagtt	gggactacag	gtgggcacca	ctgtgccagc	54840
taatttttgt	attttttgtg	gagatggggg	ttccccatgt	tggccagtct	ggtctcaaac	54900
tcctgggctc	aatcatttct	gccgcctcag	actcccagag	tgttgggatt	ataggtgtga	54960
gctactgtgc	ctggctggta	tagctttttt	tttttttttt	tttttttttg	tgacagagtc	55020
tcaactctgtc	gcccaggggtg	gagtgcagtg	gcgcgatctg	ggctcactgc	aacctctgcc	55080
tcccgggttc	aaccatttct	cctgcctcag	cctcctgagt	agctgggact	acagatg ¹ cggt	55140
gccaccatac	ccagctgatt	tttgtattct	tagtagaaac	agagtttcac	catgttggcc	55200
aggatgatct	cgatctcctg	acctcgatcc	acctgccttg	gcctcccaca	gtgctggcat	55260
tacaggcctg	agccaccatg	cctggctgtt	ggtatacttt	aaatggatgt	aattcattgc	55320
agattatacc	tcagtaaatt	tttatttagt	ttttgagact	gagttgctct	gttgcccagg	55380
ctggagtgca	gtggcacgat	ttcagctcac	tgcaacctct	gcctcctggg	tttaagcgat	55440
tttctgtct	cagcctcccc	ggtagctgga	attagaggtg	tgtgccacca	tgatcagatt	55500
atttttgtgt	ttttagtaaa	gatgggggtt	caccatgttg	gccaggctgg	tctcgaactc	55560
ctgacttcag	gtgatccacc	cgcctcggcc	tctcaaagtg	ctgggattac	aggtgtgagc	55620
caccacgcct	ggccttaatt	tttaaatact	gtaaggctta	taaagaaaag	aatattcccc	55680
ttctgtttct	ttcctctcac	gtagcaacct	tcacccttag	tttggtttca	tatacagtgt	55740
tttaactctt	tttgctattt	ttttctctta	ctttctacta	tatttcctaaa	tacaatgctt	55800
ctataatgat	tccttttttt	tttctatcag	tttttgataa	tcattgactc	cttatgggtca	55860
aagaagactt	aattcccttc	tgtcactctt	catatatata	tataactaat	atatatatgt	55920
tttagttttg	atataactaa	aaataactat	atgtatatat	aactatatat	gtataagcta	55980
tctgtatata	gttatatatg	tatatgcaca	tacatatata	gttacatatg	tatacacaca	56040
tacatatata	gttacatatg	tatatgcata	tacatatata	gttacatatc	tatatgcata	56100
tacatatata	gttacatatg	tatatgcata	cacatatata	gttacatatg	tatatgcata	56160
tacatatata	gttacatatg	tatatgtata	tgtatatata	tatatagtta	tatatgtatt	56220
agttatatca	aaataactat	gtatatatag	ctgtaaatgt	atatataaac	tatatgtata	56280
cagttatata	tgtatttgta	tatgtgtgta	tacatatata	gttttttttg	tttttttttt	56340
ttgttttttt	tgtttttttt	tggagatgga	gtcttgccct	gtcccccagg	ctggaatgca	56400
gtggtgctat	cttggcttac	cgcaacctct	gcttcccagg	ttcaagcaat	tctcctgctt	56460
cagtctcccg	agtacctggg	attacaggca	cgtggcacca	cgccaggcta	attttttgtg	56520
tcttttagtag	aaatgggggt	tcaccatggt	ggccaggctg	ttctcaaaact	cctgacctcg	56580

tgatccaccc	gcctcggcct	cccaaagtgc	tgggattaca	ggcgtgagca	ccgcgcctgg	56640
tccctataca	tatatagtta	tatataccta	tagttatttt	tagttatatc	aaaataacta	56700
tatgtatatg	taactatata	tatagtatat	atagtatata	tatactatat	agtgtgtata	56760
tatagtatat	atatactata	tagtgtgtat	atatagtata	tatatagtgt	atatatcgta	56820
tatacactat	atactatata	gtgtatatat	agtatatgta	gtatatatag	tatatatagg	56880
atagtatata	tagtatatat	agtgtatata	tactgtatat	ataggtgtac	atagtatact	56940
atatagtata	catatagtac	actgtatagt	atataatagta	tagtatatat	agtatacata	57000
gtatactata	tatagtatag	tatacatagt	atactatata	gtatatagag	tatatataca	57060
gtatactata	tagtatatag	agtatatata	cagtatacta	tatcgtgtgt	atagagtata	57120
tatacagtat	actatatcgt	gtgtatagag	tatatataca	gtatactata	tcgtgtgtat	57180
agagtatata	tacagtatac	tatatcgtgt	gtatagagta	tatatacagt	atactatatc	57240
gtgtgtatag	agtatatata	cagtatacta	tatcgtgtgt	atagagtata	tatacagtat	57300
actatatcgt	gtgtatagag	tatatataca	gtatactata	tagtgtgtat	agagtatata	57360
tacagtatac	tatatagtgt	gtatagagta	tatatacagt	atactatata	gtgtgtatag	57420
agtatatata	cagtatacta	tatagtgtgt	atagcgtata	tatacagtat	actatatagt	57480
gtgtatagcg	tatatataca	gtatactata	tagtgtgtat	agcgtatata	tacagtatac	57540
tgtatagtgt	gtatagcgta	tatatacagt	atactgtata	gtgtgtatag	cgtatatata	57600
gtatactgta	tagcgtatag	agtatatata	tagtatactg	tatagtgtat	agagtatata	57660
tatagtatac	tgtatagtat	atagagtata	tatagttaca	tatacatata	gttattttta	57720
gtaatatcaa	aaaaccatag	ctaagatttt	tatgatttag	caaaatattg	tttaccctaa	57780
agccacgtat	tatatattaa	tgatattttt	tcccttttga	gtttttcttc	cccacaggtt	57840
ttctgatctt	tttttttggt	ctctaacttt	atattatagg	tatatactct	tttctctggt	57900
taatttgctt	tttaactaga	agacctcctt	tagtagttct	tttagtgaat	atctgtgggt	57960
aaactcttag	tctttccagt	gtctgaaata	ccctcatttt	aatgtgacag	gtcacttttc	58020
tcacctcaac	atthttggaca	tacctcattt	ccaacaaaca	tttgttactt	ctgatagaga	58080
gtctgctgtc	attctcatat	ttgttccttg	tggaaatgtc	tatatataaa	gtctcagaac	58140
tatatatttt	taatatatat	acatatattt	atttatthaa	aatgtataaa	ttaagttagt	58200
agtagtagta	gtagtagtag	tagtagtagt	agtagtagtt	tttttttttt	tagatggagc	58260
ctccatctcc	cgggttcaag	tgattctcct	gcctcagcct	cctgagtagc	tgggattaca	58320
ggcacacacc	accatgcctg	gctaattttt	gtatttttgag	tggagatggg	gtttcaccat	58380

gttggccagg	ctgatcttaa	actcctgacc	tcagatggac	cacctgcttt	ggcctcccaa	58440
agcgctagga	ttacaggcat	gaaccactgc	gtctggcctt	aagttattat	tgactattga	58500
tgccctgttg	ctctatcaag	tagtaggtct	tattttattct	agaacctcgt	tttaaggatat	58560
cctgaatgtg	gctttttttt	ttttttttcc	taagacggag	tctcactctg	ttgcccaggc	58620
tggagtgcaa	tggagtgate	ttggctcacc	gcaatctctg	cctccttggg	tcaagcaatt	58680
ctcctgcctc	agcctcctga	gtagctgaga	ttacaggcgc	acgccaccac	acctggctaa	58740
tttttgtgtt	tttagtagag	acgggggtttc	accacattgg	tcaggcttgt	ctcaaactcc	58800
tgacctgtgt	atccgcccac	cttggcctcc	gaaagtctctg	ggattacagg	cgtaagccac	58860
cacgcccagc	ctaagtgtga	ttttgttttt	ttttttaact	tgaattttat	ttatacttcc	58920
tgattttata	atagttaact	tgtagtaaaa	gtaaactggg	tgtctaaata	aataagctga	58980
tttctatac	tagtttagag	tcttttagttt	tcttaaaccg	tattaagaga	tttgattaca	59040
aaagtaaaat	tgaggctttt	gtcttttttt	ttggccactc	taatttgctg	actgacttaa	59100
catttgtctt	cacagtaaac	attgtaaata	aacactgatg	tttgttgcgt	agtatgtaaa	59160
aatattattt	aggagtgggtg	agatttaaac	atttacaac	ctgtaatata	tatagtacag	59220
caacactgga	caaataacct	tttcaactta	atcagagggtg	ttctataaat	gttcatttgc	59280
atatatacag	cccatataat	acttagcata	tgtaaaagca	catctttata	ctctttcaat	59340
taaaattgat	tattattatt	attttttttt	tgagacagag	tccctctcag	tcaccaggc	59400
tggattgcag	tgggtgtgatc	agggtcact	gcaatcgggt	agctggaact	acagggatgt	59460
gccaccacgc	ccagctattt	ttgtattttt	agtggagatg	gggtttcacc	atgttggcca	59520
ggctggctct	gaactccata	cctcagggtga	tctgcccgcc	tcagcctccc	aaatagctgg	59580
gataacaggc	gtgagctact	gcacccggcc	taaaattgat	tagtttttaa	gacctcttg	59640
gtgggaacag	actttcatag	tgtttatagt	tataatagtt	tggaatcaaa	gtagtcgagt	59700
gctacttttt	tttcctctgt	acatcctatc	cctaattgct	ttactatatt	ttgtttctta	59760
tttgagagct	ttatttgctt	cataactttt	tttttttttt	tttttttttg	agacagagtt	59820
tcactctgtc	accagggtg	gagtgcagta	gtgcaatctc	ggcttactgc	aacctccgcc	59880
tcctgagttc	aagtgattct	cctgcctcag	cctccgaagt	agctgggatt	ataggccccc	59940
accaccacgt	ctgactgatt	ttttgtattt	ttagcagaga	cgaggtttca	ccatgttggc	60000
cagcctggtc	ttgaactact	gacatcaggt	gatccgtcca	cctcgatctc	ccaaagtgct	60060
gggattacag	gtgtgagcca	tcgcgcccgg	cgcttcataa	cctttttata	ttgtctttct	60120
gttccccaaa	atgaagtagt	ttcttcttaa	aaatactgtt	ttttgttttt	ttgttttttt	60180

ttttatttga	gatggaatct	ctctctgtcg	ccaggctgga	gtgcagtggc	acgatctcgg	60240
ctcacttcaa	cctctgactc	cgtggttcaa	gcaattctcc	tgccctcagcc	tcccaggtag	60300
ctgggattac	aggcacaaat	actgttttta	ataagaaaat	tagaataggg	acaggcacaa	60360
tggctcactc	ttgtaatccc	agcacttttg	gaggccaagg	tgggaggatt	gcctgagccc	60420
aggagttcga	gaccagccta	ggcaacatag	tgagtgagac	cctgtttcta	taaaaaaaaat	60480
ttttaattag	tccggtctgg	tgggtgcacaa	ctgatgtccc	agctacttgg	gagtttgagg	60540
tgggagggatg	gatcacttga	gccaggaag	attaaggcca	caatgagctg	tgatcatgcc	60600
actgcactcc	agcccaggcc	acaggtttca	aaaaagaga	aaattagaat	tatacgttag	60660
ggggaaaaaa	cctaataacc	atattcctat	ctacctagtg	accacccta	tgaagatcct	60720
ggtacatggt	ctcattgaaa	tcttattttg	aaatattttt	gaattaaaaa	aatatattat	60780
taccttggtt	ttacaaatgt	ttgcttgtct	ttatgttcag	ctacaatttt	ctaatacaca	60840
tggttttact	ttttccttgt	cagaaagtgg	agctgttcca	aaaagaaaag	acccttaac	60900
acacactagt	aattcactgc	ctcgttcaaa	aacagttatg	aaaactggat	ctgcaggcct	60960
ttcaggccac	catagagcac	ctagttacag	tggtttatcc	atggtttctg	gagtgaacaa	61020
gggatctggt	cctgctccta	ccactcataa	ggtattctgg	gacagtaact	ttaattgctg	61080
tctttttgca	aatagaaaaa	tttttaagat	actattcctg	cttaagttga	tcataagtac	61140
tttataatac	tttagagaat	ggataagttt	ccataaagtt	aaattttttt	tttttttttt	61200
tgagactctt	gttttccagg	ctggagtgca	atggcgcgat	ctcggtcat	tgcaacctcc	61260
gcctcccagg	ttcaagcgat	tctcctgcct	cagcctccca	agtagctggg	attacaggca	61320
tgcaccacca	cgtcgagcta	attttgtatt	tttagtagag	acgggggttc	tccatgttgg	61380
tcaggctggt	ctcaaactcc	cgacctcagg	tgatccgcct	gcctcagcct	cccaaagtac	61440
tgggattaca	ggcgtgagcc	accgtgccc	aagttaattt	ttttttaaat	ccacagggca	61500
acttgctata	aaactttttt	tggacctcat	tagacattta	ataccaagtt	ttgcttacct	61560
agtctgtttt	atataatgct	gtatatattca	ttttaaattt	atttttaaaa	ttatttcaat	61620
ctttgtaaat	gttttatacc	agtaacatac	ctagagtttt	gtcatgattc	taacaagggt	61680
taaaatttgt	ttttattttg	taaaaactaa	atctctggat	aaaatcctct	ataactgact	61740
taatgccaaa	tatagttctc	tagtgaatac	agttttacct	tcaggtaaat	aatatacaa	61800
tttatggatt	tttataccct	ttttcctatt	tttaaagctt	gaattctgtg	aactttaagg	61860
ttaacttatt	tatgaaaagt	gtaaatgtta	ggttgtattt	tcatattaaa	attttgtatc	61920
ctttaagggt	actccgaaaa	caaataggac	aaataaacct	tctacccta	caactgctac	61980

tcgtaagaaa	aaagacttga	agaatTTTTag	gaatgtggac	agcaaccttg	ctaaccttat	62040
aatgaatgaa	attgtggaca	agtaagTTTT	gccatctaaa	tgTTTTatTT	tatagtTTTT	62100
atattTTaat	tttacttata	aaacatgtca	ggagtgaaat	agataataaa	taccttgtct	62160
ggTTtacata	cttaatTTTT	atgatagTTT	tcaattataa	atgtagaaaa	ccattgctTT	62220
actgatttaa	ctatagTTTa	acagttaaca	ttaaaaataa	ctatatgtca	tagggcttag	62280
gcttcatctt	gtaataactg	ggccctgTTT	gtatcgtaga	actaactgag	gtcttgTTTT	62340
ttagtggAAC	agctgttaaa	tttgatgata	tagctggTca	agacttggca	aaacaagcat	62400
tgcaagaaat	tgTTattctt	ccttctctga	ggcctgaggt	aagaactTTa	tattatcatt	62460
tttctataat	accatctgTT	actgaatcca	tagtagtagt	agtagtaaag	aaatattTTga	62520
gctatactaa	aataattaat	tcatataagg	taacaataga	tttaatgTTT	taaaaaagat	62580
aaaacattaa	caattgatgt	tgagaagtta	ttatagaata	ggaaaatgga	tatgagttcc	62640
taatttccaa	tcctTTtata	TTTTtaggaa	aagtgggtag	tatgatataa	TTTTgtTTtg	62700
TTTTggagac	ggagtctcac	tcattgtgtc	gcccaggctg	gagtgcagtg	gcatgatctt	62760
agctcactgc	aacctctgcg	tcccagaatt	caagtgatTC	tcctgcctca	gcctcccaag	62820
tagctgggat	tacaggcatg	tgccactatg	cccagctaatt	TTTTgtaatt	ttagtagaga	62880
tggggTTTTg	ctatattggc	caggctggTC	tttcatacct	gatgtcaagt	tatccaccca	62940
tttcggcctc	ccagagtgtt	aggattacaa	gtgtgagtca	ctacatctgg	ccaaatTTTTg	63000
atatcaaggt	gagagagatt	taaaattaag	ataaggtaca	aaaatttagcc	tagtgtgggg	63060
gcgcacgctt	gtagtcccaa	ctactgggga	ggctgtggca	ggagaattgc	ttgaaccagg	63120
aggcagaggt	tgcagtgagc	caagatggca	ccactgcact	ccatcctggg	tgacagagcg	63180
agatgtcatc	tcaaaaacaa	aacaggccgg	gtacggtagt	tcacgcctgt	aatcccagca	63240
ctttgggagg	ctgaggcggg	cagatcacga	ggtcaagaga	ttgagaccat	cctggctaatt	63300
atggtgaaac	cctgtctTTa	ctaaaattac	aaaaattagc	tgggcttggt	ggtgtgtaac	63360
cccagctact	cgggatgctg	aggcaggaga	atcgcttgaa	cccgggaggc	ggaggttgca	63420
gtgagccgag	atcacgccac	tgactccag	cctggcgaca	gagcgagact	ccgtctcaaa	63480
aaaataaata	aaaattttaa	aagataaata	cataaaaata	aataaataat	attaagagaa	63540
ggaaatcagg	caggtagtgg	cccctgacac	aatgagTTTT	cccagaattg	gattgcttgg	63600
aaatgccgct	caaagagtgt	ggtaaactcc	atcgaaggct	aaataccaac	gtgacagtga	63660
taataaacaa	gtactTTtatg	ggaaagTTTT	TTTTTTaatt	TTTTTTaaaa	agagagaaat	63720
tgtactggag	aaaagaggaa	ttcaggtaga	ataattcttt	TTTTTTTTTT	TTTTtgagatg	63780

gaat t t t t g c t	c c t g t t a c c c	a g g c t g g a g t	g c a g t g g c t t	g a t g t t g g c t	c a c t g c a a c c	63840
t c t g c c t t t g	g g g t t c a a g t	g a t t c t c a t g	c c t c a t c c t c	c t g a g t a g c t	g g g t c t a t a g	63900
g c a t g c a c c c	c a a c a c c t g g	c t a a t t t t t g	t g t t t t t a g t	a g a g a t a g g g	t t t t a c c a t t	63960
t t g g c c a g g c	t g g t c t c a a a	c t c c t g g c t t	c a g a t g a t c c	g c c c a c c t t g	g c t t c c c a a a	64020
g t g c t g g g a t	t a c a g g c a g a	g g c c a c t g t g	c c t g g c a t g a	a g a a c a a t t t	t t a a a a g a a t	64080
g a c t t t t a a g	g a t a t t a a g t	c a t c a a a g t a	g a t a g a g c c a	t t a a a t g a t g	g g t a g a a a c c	64140
t a a t c t t c c a	t c c c a t a t t t	t a t g a c t t a t	t a a g g a a g a t	a g g c a a t c t t	g c a g t t g a t a	64200
a t a a a t a t t t	g g c t t t c c a t	a c t t a g c a c c	g t t t t g a a t t	t t t c c a g t t c	a c a g a t g g t a	64260
t a t t t a g t g c	t g t t c c a a t a	a t a t t g c a a a	a a a t a a a t c t	g a a g a c t c a c	t t c t g g g t c a	64320
c a g t t a t t t c	a c t a t t a a c a	t t a a a a t c t t	a c g g a c c t a c	c t g c a a c a t g	t a g t g g a a a t	64380
a a g t t g t g t g	g c a c g t t g t g	g t g g g t g c a t	t a t t a a a t a a	a t g t g c a a a g	g t t t t a t g g c	64440
t g t c t c a a t t	t t t c c a t a g t	c t t g a g t c a t	t c a c a t a c t g	c c t t t a t g g t	g t t t a t c a c t	64500
t g t a a g t g a g	t a c c a a c t g t	a c t a g a a c t t	a c t g a a t a t t	t t t c a a g t c t	a t t t t a c t t a	64560
t t a c t c a t t c	t a a g t a c a t a	a a a t c a c t g g	t c t g a t a t g c	a a g t t a g g t t	t t c g c t a a t a	64620
c c c a g t a a a a	t a a t t a g a t a	a t t a t t a a a a	t a g a a a g c a t	t t g t c a t t a t	g t g a a c t g g a	64680
t t t a t g t c a t	g t a c c a c c t c	a g g t c a c a c t	t t g g g a a a c g	t a g c t t a c g t	a t t g a g c t a a	64740
c a t t t t g c t g	g a g t t a t a t a	a g a t t a g t a a	a t g c t a t a g a	c t a a a t a t t g	a a a a a a a a g c	64800
t t g c t t a a a g	a g g a t t t a c a	t a a t t g t a g a	c c t g a a a a g g	a t c t t a g c a g	t c t t c t g a t c	64860
g a t t c a t t g a	t a a g a a a a c t	a g t c c t a a g a	g c t a a a a t g a	c t t g c t c a a g	a t t t t a g t t a	64920
a g t g g t a g a c	c t a g g a a a t a	c c t t t c t a a g	a t t a g a g c t g	c c t g g t t a g t	g c t a t a g c t g	64980
c c a c c t a a t t	c t t a c a t t t a	t t t a t t a a g g	g a c a t g c t t t	t t a c g g t a g a	a a a a t a t a t c	65040
t a a a a t t g t g	a a g g t a a t a c	c a a g g a a a t a	t t t a g g c t a t	g c a a g a a a a g	t c a t t c a g c t	65100
t t t g c t t t a c	c c a g g c c t a t	a g a g a a c a t a	t c t a a a a t a a	t a a g t a a c a t	g t g a a t t t g a	65160
a t t t a g g c a t	a t t t t c a t g g	t a g t g a c a t a	a t g t g c a a t t	a a g a a g c t a t	a a t a a t g t a a	65220
a a a t g c c a g t	t g c a a g t t a t	t t c a a a t a a a	g t t g a a t a a c	c t g a c c t c a g	g a g g g g c a g a	65280
a a c c t t a g t a	a a c t t a g c a a	a c t g a g a c c t	a a g g a a g a g c	a a g a t c t g t g	g t a g a g c c a g	65340
g t c t a a a g g t	t t a t a t c c t g	g t a c a a a c t t	g t c t a t a c a t	c a c c a c a c c t	g c c a a a t g g a	65400
a c t t t t t t t c	c t t g t a g t t c	a g a t t t c c a a	g a a a g g g a a t	a t t a t t g t c c	t a g c t c a g c g	65460
t t t c a c a a t a	g a c t c a c a a a	a t g a a t c t t g	t t a a g c c t a t	g a a t g g g t a a	g a c t t a t c c c	65520
a t a a c a c g c c	t c t a t t g t g t	a a a a a a t c a	g c t t t a t t c c	t a a t t c c c t g	t a g t a g g a a g	65580

cttgcataga	ccacaataga	gacattgtta	gattatatca	agaaggtgag	tggcaggcat	65640
tcttgagtat	actggaatth	ttatttcatt	gtagaaacac	agaatcataa	ttagtcatct	65700
aaaactggth	ttaaggagat	aatactaaag	atactcactt	gaggtagatt	gttctcagaa	65760
ttacttaaaa	taagtgcctt	aaatcttgct	cattcatgaa	gccatagccc	tggaaggaga	65820
tattgaaaaa	cttctccatc	ttccttcaat	cccataggac	catgttttta	attgtagttt	65880
aacattaaac	ataatttttg	aaagtagcca	atattaatct	ttgacaaaaa	actcaaacta	65940
aatagcagaa	aatagaacat	gccctctcct	ctcgcccaaa	agtaagacta	catttttaag	66000
tatatcacag	tgatctgtct	aaatatthtt	ctcaaataat	ttactaagaa	agcgtatat	66060
cttctgtcct	tatttcatag	atctcttctg	tatcctgtgc	agcttgtcta	tcatttcogag	66120
cttttatatt	tcttatthtg	ttctcttgct	gcttaaatgt	gattctttga	gatttataga	66180
agaatgtgat	tgthctccag	ttttaaaaag	ctcttctcta	tctactaaaa	tttccatgaa	66240
tttctttctt	tctttthttt	tttttgattg	agacagtctc	gctgtgttct	ccagcgtggc	66300
cttaggcctta	agtgatcctc	ctgtctcagc	ctcccaagta	gctgagatta	cagaaatttc	66360
catgaatttc	caaatatgat	gttatctthc	atataagctt	aaacataagc	ctttcttctt	66420
tcaaaaatgt	tagaacttac	ctctthaaat	aatcttataa	agccatttct	taattthtgt	66480
tacctthtgc	aatctattaa	cttgcataat	atagtaaagt	aacccttagc	tccatcatac	66540
cacacacctt	ctactcctgt	cctcccactt	tttctcacc	aggctggagt	acagtgaata	66600
tttagtgctt	ttaaatatta	ttcattacca	agccaagtag	cgtactatth	tttcttgta	66660
cagcctthtt	ttcccctaaa	taattgtgtt	atthtttcat	ttgcttcatt	ttcatcttat	66720
cgattgataa	ttcttctgtt	ccatttcagt	ctaatttcca	catggccaaa	ccaatctaaa	66780
aacctgttaa	ctcaggtctt	ttttthtttt	ttttthccca	gaatacctth	tcctggaacc	66840
accttcttct	tttcattcga	gaccaaagtc	tttccatgcc	tgtgtcaaag	tggttgttct	66900
ctaactctgt	ttcactacca	tcataatttc	ctthtctctt	tttctthtct	ttattatgta	66960
tctcctgtct	tttaaatthc	atattthtcca	ctthcatgat	ttattthtct	atthtgatgg	67020
ctacatctth	cagtaactga	gaaaagggtg	tgtgagggaa	tattthgaga	tgctgtagtt	67080
ttgaaaatat	ctthattctg	ctgtcttctc	attaataatt	tgactggatg	tcaagattta	67140
agttggaaac	tattthccat	gagtattthg	aaggcattat	ctattgtctt	ctagcttcca	67200
gtgctgctth	gtattagaga	cagaagatca	ggaatcagga	tagcattgga	cttcttattt	67260
ctattthccca	aagctthttg	gatcttctct	ttattcttat	cttctgatat	tttataatga	67320
tatgcttgat	gcattthact	tttctthtct	tcttccctct	tgtacatttc	tttctthttt	67380

taggtggggg	caggggggaa	ggagtctcac	tcttctcacc	caggctagag	tgcaagtggg	67440
caacctcagc	tcaactgcaac	ctccacctct	gggggttcaag	cgattctcct	gcctcagcct	67500
cccgagtagc	tgggattaca	ggcatctgcc	accacgcca	gctaattttt	gtatttttag	67560
tagagacggg	gtttcaggct	ggtcttgaac	tcctgacctc	aggtgatccg	ccctcctcag	67620
cctcccaaag	tgctgggatt	agaggcatga	gccactgcgc	ccagcccctc	ttggaccttt	67680
ctagatattc	atccttcaat	tctgagaaag	gttctataat	tctttaattt	aaaaaaagt	67740
tttttgtttc	atctttctca	cttttttttt	tttttttttt	ttttaagaca	cagtcttgct	67800
ctgtcaccga	ggctggagtg	cagtggtgcg	atctcagctc	actgcgccct	tcgcctcctg	67860
ggttcgagtg	attctcctgc	ctcagcctcc	caagcagctg	ggattacagg	tgcccaccac	67920
caccatgccc	ggctaattgt	tgtattttta	gtagagacgg	agtttcacca	tgttgccag	67980
gctggtttca	aactcctgac	ctcaggtgat	ccacctgctt	tggcctcca	gagtgtggg	68040
attacaggca	tgagccaccg	tgcccagccc	tattctgatt	tcatagatgc	agtgtctttt	68100
atgtctctga	gaggttttgt	caagttttct	tttgtttact	ttattatctg	tttcttcaag	68160
attcctcctc	ttttcttttt	gccttaagct	ttttcatatt	ggaagcttct	ctcaaagtgc	68220
tggttattct	ggttgtccac	ttatattacg	tggaacaata	aaaaagttga	ttaggactct	68280
gtgctgtaag	taagtaaatt	gttgatagtg	agagccatt	aatcacagga	tgatcaagca	68340
gcaagcctac	tgtttatttg	ggtttctcca	aatggctata	gctgtcagtc	tttttttct	68400
ggggttatth	ggcttctcta	gaaaagaatt	ttctagtctc	ttgtttggag	aatacaagct	68460
tgactattgg	tgttccagaa	aatgggtggg	aggagacgac	tttgtttct	gttaggttga	68520
gccatgtgaa	attactaggt	tttgttttgt	tttctgtttt	gttttgagac	agagtctcac	68580
tctgttgccc	aggctggagt	gccttagcat	gatcttggt	cactgcaacc	tctgcctccc	68640
aggttcaagc	aattctcatg	cctcagcctc	ccaagtagct	gggattatag	gcatgcacca	68700
ccacgcctgg	ctaatttttg	tatttttagt	agagacgggg	ttttgccatg	ttgaccagtc	68760
tagtctcgaa	ctcctgacct	caggtgatcc	acctgccttg	gcctcccaa	gtgctgggat	68820
tacaggtagt	agccaccaca	cctgggctga	aattactgtt	tttataggtc	aaaaacagtt	68880
gagggatagg	catggtggct	cccacctata	atcccagcac	tttcagaggc	caaggcagga	68940
agattgcttg	agtccaggag	ttcgacacca	gcctgggcaa	tgtagtgaaa	ccccatctat	69000
atttaaaaaa	aaaaaataga	tgaatatctt	tgtttgaga	taacatgatt	gcatatgtag	69060
aaaatcccaa	agaaccaaca	aaaagagctc	ctagaactaa	taagtgatta	tgacaagggtg	69120
tagaatacaa	agttaatatata	caaagtcaca	ttgctttttt	atctaccagc	agtgaacaac	69180

tggaatttga	aattaaaaca	caatactgct	gggtgcagtg	gctcacacct	gtaatcccag	69240
cacttttgga	ggccgaggtg	ggcagatcac	ctgatgtcag	gggttcaaga	ccagcctgac	69300
caatatggtg	aaaccccatc	tctactaaaa	atacaaaaat	tagctgggca	tggtggcggg	69360
tgcctgtaat	ctcagctact	caggaggctg	aggcaggaga	atcgcttgaa	cccaggaggc	69420
agaggttgca	gtgagccaag	atcgcgccat	tgcactccag	cctgggggac	agagcaagac	69480
tccgtctcaa	aaaaaaaaac	acaatacctt	tcatattaac	actaataaaa	tgaaatatgt	69540
gtagttctaa	caaagtttgt	tgtagaagat	ctatatgaga	agaattatag	cactcatgaa	69600
agaaatcaaa	gatctaagta	aactgagaga	tattccatgt	aaatggacag	ggagactaaa	69660
tattattgag	atgtcagttc	ttcccaagtt	catatatcga	ttcagtcgag	tcccagtcaa	69720
accccagcca	gttattttgt	ggatactggc	aaactaaagt	ttatatgaaa	aggcaaaaga	69780
cctagaacag	ccaacacagt	attgaagaag	aaaaaagtca	gaggactgaa	actacccaat	69840
ttcaagactt	actgtaaagc	tacattaatc	aagacagcat	gtcattggca	aaagaataga	69900
caaataaatc	agtagaattg	gacagagagc	ctagaaatca	acccacacag	ataaagtcaa	69960
ctgatctttg	gcaaagggac	aaagacaatt	cagtgaggaa	aagatagcct	tttcaacaaa	70020
tggtatagga	caactggaca	tccacatgca	aaaaagttaa	tctagacaca	gacctgacaa	70080
ctttcacaaa	aataaatgga	tcatagacct	aaatgtaaca	tgcgaaactga	aacttctaga	70140
ggataacata	ggagaaaatc	taggataaca	tgagaaaaat	ttttggtttg	gcagtgactt	70200
cttaggtaca	ataccaaaac	atgatccttg	aaaaaaaaaa	tcagtatggt	gaactttggt	70260
taaattataa	acttctgctc	tgtgtaagat	gctgttagga	gaatgaaaag	acatgcagca	70320
gagtgggaga	ctttacaaaa	ttcattatct	gatgaaggac	cagtatccaa	aatatacaaa	70380
gaacttttaa	aactcaacaa	taagaaaata	tacaaccag	ttaataaatg	ggcaaaatat	70440
ctgaacagac	acctcaccaa	ggaagataga	tagatgacaa	caagcatatg	aatatatgct	70500
caacatcatg	tcgttaggga	aattgcgcat	taaaacaaca	acaagatacc	ctgccatccc	70560
tattagaatg	gctgaaatct	aaaacactga	caacaccaaa	ttctggcagg	gatgtggagc	70620
agcagaaact	ttaattcatt	gctgatggaa	atctaaaatg	gtagaaccat	tttggaaggt	70680
agttggacag	ttttttacag	aactaaagac	agtttgacag	tttcttacia	aactcttacc	70740
atatggtcca	gcagtcttac	tccttagtat	ttacccaaat	aagtttaaaa	tgtacatcca	70800
ataaaaaaaaa	ctgcacatga	atatttctag	cagcattatt	catagttgcc	aaaacttgga	70860
agcagtcaag	gcatccataa	gtaggggaat	ggataaacag	actttgggtat	atcatgtaat	70920
ggagtattat	tcagcaataa	aaagaaatga	tctatcaagc	cacaaaaata	tatggaggaa	70980

ccataaatgc	atattgctaa	atgaaagaag	ccagtctgaa	gaggctacac	tataggattc	71040
tgactatatg	atgttttgga	aaaggcaaaa	ctatggaaac	agtaaataga	tcagtgggtg	71100
ccaagggaga	caggagagaga	tgaatagggtg	gagcacagtg	gattttttaag	gcagtgaaac	71160
tgttctttat	gataatccaa	tgggtggatac	atgtcattat	acctttgtca	aaaccacag	71220
aatataaaac	ataagagtga	accctaattgt	aaaatatgga	cttcagttaa	taataatata	71280
tgaatatattt	ttcattagtt	ctaacaagtg	tactacacta	atacaagata	ttcagagtag	71340
gggaaattgg	aaaggaatga	gaggttatat	gggaactctg	tactttctgc	tcaattttct	71400
gtaaacctaa	aatcactaaa	aaaaaagttt	atttttattt	ttattttttt	ttaattttta	71460
attttttgag	atggagtttc	actcttggtg	cccaggctgg	aatacaatgg	cacgatctcg	71520
gctcactgaa	acctccgcct	ccagggttca	agcgattctc	atgcctcaac	ctcccagta	71580
gctgagatta	caggcatgtg	ccaccacgcc	cagctaattt	tgcagtttta	gtagagacag	71640
ggtttctcca	tgttggtcag	gctggtcttg	aactcctgat	ctcaggggat	ctgcccgcct	71700
cggcctccca	aagtgcgtg	attataggcg	tgagcagctg	cgcccagcag	gttttttttt	71760
ttaagttgga	tattagccat	ttcatatgat	tcaacttaaa	agtacataca	ccttcactgt	71820
tattaaagtg	taggtgagat	gttttcagtc	tggagctcta	cccttgattc	ctgccatgcc	71880
tagtgtcctt	gaatctggag	actctgactt	atctcttttag	agaacgaaac	tcctgccttc	71940
tgctttgggtg	gtgattggta	cctgcttgac	tgcctttggt	ggggagttcc	tcataccaac	72000
tttcaatcag	ttcctgtgtt	tgtttgtttg	tttgtttttg	agacggagtt	ttgctctgtc	72060
gcccgggttg	gagtgcagtg	gcgcaatctt	ggctcaatac	acgctccgcc	tcccgggttc	72120
acgccatttt	cctgcctcag	cctcctgagt	agctgggacg	acaggcaccc	gccaccacgc	72180
ccggctaatt	tttttgattt	tttagtagag	acgggggttc	atcgtgttag	ccaggatggt	72240
ctcgatctcc	tgacctcggtg	atccgcccgc	cttggcctcc	caaagtgtg	ggattacagg	72300
cgtgacgact	gtaagccacc	gtgcccggcc	agttcctctg	tttttgactg	cctgccttac	72360
tactgctttc	tgtggtgcct	gataacatcc	aattcctgaa	ccttcctggg	atttttgttc	72420
acgtcagcgt	gcttcttgca	tttaggtatc	ccttcacaag	taggcattta	ggttttaagc	72480
tctgctaagt	gatttaccac	actttatcag	ttctccattt	tgtggaattc	attgtgttaa	72540
tctccttttc	tatttttggt	tggaagattc	atatattttt	tattcattta	gtattttttg	72600
gtgggatttc	taagtagaga	aacatgccta	tgttcaatat	gtcttgttta	agcagtctgc	72660
tttcattttt	accaccgagg	agttgggttt	atctcttttt	tttttttttt	tttttttttt	72720
ttttgagaca	gggtctcact	ttgtcagcca	ggctggagta	cagtggcaca	atcaaagctc	72780

actgcaccct	tggcctccca	agctcaagtg	atcctcccac	ctcagcctcc	tgaatagctg	72840
ggagctcagg	tgatccagca	caccagccta	atTTTTTTTT	TTTTtaattt	tttgtagaag	72900
cagagttccc	ctatgttgcc	caagctggtc	ttgaactctt	gggctcatgt	aatcctcctg	72960
tctttgcccc	ctaaagtctg	ggattacagg	tgtgagccac	cacaccagc	tgTTTTtaat	73020
tattacatta	atTTtatgatt	atgtgtttcc	TTtaaagcta	tgggcagctc	tgTTTgggaa	73080
gatgctactg	aaaaaaggat	gctTTTTtaga	tggcaaagag	tacttaaaat	gtctctagaa	73140
tcatagtgtg	aaactaaagt	atatatTTTT	tagttgttca	cagggcttag	agctcctgcc	73200
agagggctgt	tactctttgg	tccacctggg	aatgggaaga	caatgctggg	aagggttctc	73260
ttcaaatttg	agTTTTctgt	tgagatatTT	gggataatat	gaaaaaaaga	aactTTtatct	73320
tgTccttgag	tctattatTT	acgacttgct	TTTTgctatt	gtacactTTT	gTTTTTTTTg	73380
TTTgtttgtt	TTgttttgtt	TTgagtgatc	Tgggctcacc	gcaacctccg	cctcctgggt	73440
tcaagcgatt	ttcatgcctc	agcctcccaa	gtagctggga	ttacaggcac	acgccaccac	73500
acctgtctag	gTTTTttatt	TTTggcagag	acaggatttc	gccatgttgg	ccaggctggg	73560
gtcgaactcc	tgacctcaag	cagTcctcct	acctgggcct	cctaaagtga	tgggattaca	73620
ggcatgagcc	actgcacctg	gctgatacac	TTTTaagttt	ttcagctact	TTTcaatgta	73680
gaagtagatg	gaaaaccatg	tacgttatct	tcagtagtgt	gTTTTtggtt	ggttaaattt	73740
gacagtatga	TTgtcattat	TTTTtgtaaa	TTaaattttt	acctggaaga	gcttacctta	73800
ctatattgag	tatctttcta	acctctgatt	TTTgcttcta	ctatcataat	aactTTattt	73860
aagtaatcag	tatgttatag	ctTTTTTTTT	TTTTtaagta	TTctTTTgcc	agaagTTTTt	73920
atcaggctct	ggatacctct	ttcctctgca	tagTcctcct	ggatggaaga	aacaaagagg	73980
gaaagagtaa	cttttcctta	gatgtttgtc	TTTctcaaag	cagttatctt	tgtatatcta	74040
agaagagagg	agaataacac	tgtctctctt	TTTTTTTaaa	tctctctcta	ctcattctct	74100
ctcaggagga	gagtagaaag	aagcacagct	cttcctataa	cctgtcctta	ttactgagaa	74160
aggaacacat	tgattgccat	gtattgggga	TTgtattata	ccttacattt	TTatTTTTat	74220
TTTTtattaa	TTTTTTTTTT	tagacagtct	tactctatca	cccaagctgg	agtgcagtga	74280
Tgtgatctcg	gctcactacc	atctcttcct	tctgggttca	agcaattctg	ccacgtcagc	74340
ttcctgagta	gctgggatta	caggcatgcg	ccaccacgcc	tgactaattt	TTgtatTTTT	74400
attagagatg	ggatttcacc	atgttgGCCa	ggctggTctc	aaactcttgg	cctcaagtga	74460
Tgcgcctgcc	tgggcctccg	aaagtgcTgg	gattacaggc	atgagccacc	acacctggcc	74520
tcatagctta	cattTTTTtaga	gaatctTTTT	tagtacttaa	atcggtaaat	atggTtatct	74580

tttaaagtga	atatattgaa	ctaatttaaat	atttgctcctt	gtgatttttta	aaggctaaag	74640
cagtagctgc	agaatcgaat	gcaaccttct	ttaatatataag	tgctgcaagt	ttaacttcaa	74700
aatacgtgag	tgctctgttt	ccaatattgt	cgtatttttaa	gttactgtct	aaatgttact	74760
gtgttaactg	taaatggtaa	tatttcatga	aaatattttt	ctaggagcct	atctattgta	74820
tctattattt	acatatgatg	aatatctatc	ttcagagtag	aaagtatatgt	acatttgtgt	74880
tgtcaaatac	tgtattagtt	tactggggcc	atgtaataaa	ataccgtaaa	ctggctggct	74940
taaacatcag	caatttattg	tctcacagtt	gtggaaggta	gaagtctaag	atcagtcaaa	75000
atgttggcag	ggttgcttcc	tcctgagggc	tgtgagggaa	aatgtatggt	gtgtgcctct	75060
ctcctggcct	ctggtggtt	gctggcaatc	ttttgtattc	cttggttgt	agatgcatcc	75120
ctcctatctc	tgtctttatc	tttatgtggc	attctccctg	tgtctgtcac	catgtccaaa	75180
tttccctgt	ggattaggac	ccaccctaata	gatctcaatt	taagtttgtc	atcagcaaca	75240
attctatgtc	caaataatgt	cacattcata	ggtactaggg	ataggacttc	aacacgtttt	75300
tggagaacac	agttcaaccc	attaacaaat	actatcactt	tccacttaag	cttcaagtaa	75360
agtggatttt	atctcaagga	gccaccagat	aggaacacag	atctgatggc	ataaactgag	75420
tatttctggc	cttctgattc	tgatctaaaa	tatgacagaa	gattttccct	gttttaattt	75480
tttttttttt	tttataatag	agacacagtc	tcgctatggt	gccaggctac	tttcaaattc	75540
ctgggcttaa	gcaatcctcc	tgctcgggac	tctcaaagtg	ctggaattac	aagtgtgagc	75600
caccacacct	ggccagtaga	ttttccctgc	tttcttttga	ttgtttataa	ttttgttttc	75660
tttttccatt	cacctctgc	tgacctata	gtattattca	aagaagtgtt	cagtctagtt	75720
ttggggtagg	gcaagcataa	ctacagtgt	taaagagagt	aatttgtctg	gtgtgcagaa	75780
ctagtgtgta	aatataactg	gtgcattgca	aaactgtgaa	gtagtttctg	tcaaacctta	75840
cactgctttg	tctttctccc	tctctccctt	tctcttggtt	gccccctccc	cctcccaatg	75900
ataccttagt	ctctgcttgc	ctttatcaaa	acctttatga	ttggccgggc	acagtggccc	75960
atgcctataa	tcctagcatt	ttgggaggct	gaggcaggaa	gatcacctga	gcccagaagt	76020
tggagaccag	cctaggcaac	atggtgaaac	cccatctctt	ccaaaaataa	aaaatagagc	76080
caggtgtggt	ggcatgcacc	tgtagtccca	gctactcagg	aagctgaggc	gagaggatct	76140
cttgagcctg	agaggttgag	gctgcagtga	actgtgatca	tgccactgca	ctccagcctg	76200
agtgacagtg	tgagaccctg	tctcaaagaa	acaaaacaag	ggggggcatg	gtggctaact	76260
cctgtaatcc	cagtactttg	ggagactgag	gcaggaggat	tgcttgaggc	caggagttca	76320
ataccaccct	gggcaacata	gtgagacccc	catctctaca	aaaaataaaa	aatttagctg	76380

gacatgccag	cgaatacgtg	gtcccagcaa	atcaggaggc	tgaggtggga	ggatcacttg	76440
agtcaaggag	gttgaggttg	cagtgagcca	cgatcatgcc	aatgcattcc	aacctgggcg	76500
tcagagcaag	accacgtctc	aaaaacaaaa	caaaacaaac	ttttatgggt	gaaagtgttt	76560
tggaacaacat	acttaaacgt	aaatgtgaat	ctctgatgaa	agaacatggt	acctgtaaaa	76620
gtttgaagtg	tcagcatttg	ttgcaccgaa	atccagaggt	gaggccaggt	gtggtggctc	76680
acgcctgtaa	tcccagcact	ttgggaagcc	aaggctggca	gatcacctga	ggtcaggagt	76740
tcaaaaccag	cctggccaac	attgcaaaac	cccgtctcga	ctaaaaatac	aaaaattagc	76800
cagacttggg	ggtgtgcgcc	tgtaatccca	gctattcggg	aggctgagac	acgagaattg	76860
ctcgaaccca	ggaggcagag	gttgacgtga	gccgagatgg	caccactgca	ctccagcctg	76920
ggcaactgag	tgagactctg	tctcaaaaaa	aaaaaaaaaa	aaaaatccag	aggtgaatcc	76980
agaggtgatc	accacatgat	atccagatag	cctctttcca	tgagaggctc	aaaggataat	77040
tttactgtct	acagttttgc	agcgagagaa	acttgatttt	atcagtacac	caagagcaga	77100
tctatgtctt	cggaacagac	atgagatcag	aattgtctag	ctgctatgaa	cagcatgttc	77160
tctccctgta	cctatagaca	tgtatgggaa	acttatttgt	aaggttgtat	aatgagcagt	77220
gagttaaagc	aaacttgaca	tgttgaccat	agttgttatg	gcattggact	aaagtagcct	77280
ccatcactat	gatagagata	gccttggtgt	tggaccata	gtttttgaag	tctgtttgct	77340
aagactccct	cttccttagg	cagttcttca	aatattatta	tgctttttcc	tgtcatggga	77400
ttcctcttca	tggagtttcc	tttctcagga	acactatctt	ctcctgggta	atltgtactt	77460
attcttcaga	taccagttga	aatgttacaa	ccttccctga	ccctccaaac	tattccccgt	77520
tattattctc	ctagcaccac	ttgcacctca	ttttcatact	cactggaggt	gcaattcata	77580
ttcattgata	ggattatfff	aattgtatct	gatatcactg	tcacctccac	tagaagatgg	77640
tctccatgcy	ggcagagact	atcaccatgt	gttcttcact	tcaattttca	gtagttggct	77700
gtgagtaggt	attgaataaa	tatttgtgga	gtaatcataa	tgaggtatag	atattattct	77760
catagctcta	ttttattaat	tagggaatta	cagaattcag	tgatctgctc	aggatctcat	77820
aaccaggaag	tgggaaacta	ggatttgagc	tccagtgagt	gtggcctttc	attaaaaata	77880
ttacagcaac	cattctcttt	tttttttttt	ttttttttga	gatggagtct	tgctctggag	77940
tgcaagtggcg	tgatctcggc	tcactgcagc	ctctgcctcc	cgggttcaag	tgatccccct	78000
gccttagcct	gctgagtagc	tgggactaca	ggcaccgcgc	acgacacttg	gctaattttt	78060
gtatttttagt	agagacgggg	tttcaccatg	ttggccagga	tggtctcgat	ctcctgacct	78120
tgtgatccgc	ccgcatctac	ctcccaaagt	gctggattat	aggcgtaagc	caccacaccc	78180

ggcctacagc	aaccattctc	ttttatccat	acttttttca	agagtactgt	ttcatcttca	78240
tgttttcaga	aacaacatag	cattcatgat	cttaaccccc	aattctgata	ctgcctgaat	78300
atcttgaagt	aagtttactt	ttaagaaagt	tgaggctagg	tgtggtggct	catgcctgta	78360
atcccagcac	tttgggaggc	caaggcaggt	ggatcacttg	agctcaggag	ttcaagacca	78420
ggctgggcaa	catggcgaaa	ccctgtctct	accagaaata	caaaaaatta	gtcgggcgtg	78480
gtggcgtgtg	cctgtggtcc	cagccacttg	ggagactgaa	gtgggaggat	ttcttgagct	78540
tgggaggtgg	agtttgctgt	gagccgagag	atcatgccac	tgtactccag	cctgggtggc	78600
agagtgagat	cccatctcaa	aaaaagaaaa	gaaaaagaaa	attgaaatgt	ctagtctatc	78660
attttgtcag	ttctatctaa	tacaattttt	tccttatgtc	taactgaaat	ctgctttttc	78720
taatttttac	atacttgatt	taacaaaact	caatcttttt	ttttttttat	gagacagcct	78780
ttcaaataata	taggaactta	atgttatatc	tgttccctc	cagtccccag	aatagttact	78840
attttagttg	tctttctatg	gtctcatgcc	agtttgtcag	tatgcctaga	taagaactga	78900
atattttacc	tcagatgtga	cctgactttg	aagacttaaa	aaggaagcat	tgtgccaggc	78960
gcagtggctc	acacctgtga	tcccagcact	ttgggaggcc	gaggtgggca	ggtcaggagt	79020
ttgagaccag	cctgattaac	atggagaaac	cccatgtttc	tctactaaat	acaaatctct	79080
actaaatact	aaatactaaa	tctctactaa	aaatacaaaa	gatgagctgg	gcattgtggc	79140
acatgcctgt	aatcccagct	actcgggagg	ctgaggcagg	agaatcactt	gagcccagga	79200
ggcggagggt	gcggtgagcc	gagattgcgc	cattccactc	cagcctgggc	aacaaaagtg	79260
aaactccatc	tcaaaaaaaa	aaagcagcag	cattgtgtaa	tattatgtag	atgttggtgc	79320
tcatgatcta	tcctgagaaa	gcttttgggg	ggaactgcat	catagtcatg	gacaacattt	79380
gtgttattaa	aatatctaga	ttattttcca	caaaaaatca	gttacatatg	tatcttaaca	79440
tgttgtatta	ttgtttaacc	ttgtttattg	aataactaac	atgtagaaaa	gtattttatag	79500
cataagtata	cagctgtata	cagctcagtg	gattaccaca	aagcgaatat	actttcataa	79560
tcaccaccca	ggtcaagaaa	taaattgtta	cctgtggccc	taaaatccct	ccaggcactc	79620
caccatcttt	atccactcac	tcctctccct	caaaaccact	agactactaa	catcatagac	79680
aaagctagca	tgcctttgaa	ctttatataa	atctaataat	gtaggatttt	gtgtgtatgt	79740
gtatttggct	tctttcatca	gcattgtatt	tgtgagattt	atccagattg	ttgcaagtag	79800
ttgtagttgt	gcttttttac	acagatttaa	tttttatatt	tttcttatat	gttcgaacag	79860
ttaacctgct	tatctattat	taaaaaaaaa	aaaacgaaca	ttcacatagt	tcttaccagt	79920
ttacagtgtt	ttttccacac	cgtcttcaaa	atgtaaagtt	tggtcttcaa	tacatcagta	79980

tgctgctaga	tttaaatact	agggaaaaaa	aaatcagaga	agttaataat	attaactgtc	80040
acctccacta	gaagatggtc	tccatgtgga	cagtaatatt	tctcttgat	tatctgtgct	80100
aagtaaaatc	ttctgtaagt	ttctttaaat	attttaataa	atcatagtac	ttaaaatggt	80160
ctcaatattc	taaagtagtt	aaaagtaact	ataaaatagt	acctgttttt	ctgatcacat	80220
tttacttcct	atgtgaaatt	ttacaagtcg	ttactctatt	tattttattga	tttatttttt	80280
aagacagggg	ctgttctgtc	gcccaggctg	gagtacagtg	gcgtgatcat	ggctcactgc	80340
agcctcagcc	tcttgggctc	aagtgatcct	cccaccttag	catcccaagt	agctgggact	80400
ataggcacat	gccaccatgc	ccagctaatt	ttaaaaaatt	ctggggggcc	gaatgcgggtg	80460
gctcacacct	gtaatcccag	cactttggga	ggccgaggca	ggcgaatcac	aagggtcagga	80520
gttcgagacc	agcctggcta	acatggtgaa	accctgtctc	tactaaaaat	acaaaaaatt	80580
agcgggggtg	ggtggcaggc	gcctgtaatc	ccacttactc	aggaggctga	ggcaggagag	80640
ttgcttgaac	ctgggaggca	gagattgcag	tgagccgaga	ctccatctca	aaaaaaaaaa	80700
attttttttt	gtagtgacaa	ggtgtcactg	tggtgccagg	gctggtctca	aacttctggg	80760
ctcaagtgat	cctcccattt	cggcctccca	aagtgctagg	atcacaggca	tgagtactgt	80820
tgcttggctt	tcaagttggt	attaaagcat	gtttaccac	attatgcaca	tggtataatg	80880
gaaagtattg	ttgtggaagt	taggagatag	ggattctagc	ctagcttttt	atttttttgg	80940
gacaaggctt	cactttttcg	cccaggccg	aagtgcagtt	gtgcgatctc	ggctcactgc	81000
aacctccaac	tctcaggttc	aagcaattct	cccacgtcag	cctcccgagt	agctgggatt	81060
acaggcatgc	gccaccacgc	ccggctaatt	tttgtagt	tagtagacac	agggtttcac	81120
catgttggcc	aggctggtct	tgaactcccg	acctcaggtg	atccaccac	cttggcctcc	81180
caaagtgtg	ggattacagg	catgagccac	cgcaccggc	ctctagcgta	acttttacat	81240
cctgaactga	ccttaagaaa	gtataacttt	aggcctgttt	catctgtaaa	atgttaatgt	81300
cataggagat	gatcttttga	gatttctttc	agctctgata	attttgtgtg	tgtgtgtgtg	81360
tgtgtgtgtg	tgtgagatgg	agtcttgctc	tgtcgccgg	gctggagtgc	agtggtaacca	81420
tctcggtca	ctgcaagctc	cacctcctgg	gttcacgcca	ttctcctgcc	tcagcctccc	81480
gagtagctgg	gactacaggc	gcctgccacc	acgcctggct	aattttttgt	atttttagta	81540
gagacgaggt	ttcaccgtgt	tagccaggat	ggtctcgatc	tcctgacctc	gtgatccgcc	81600
cgcctcagcc	tcccaaagtg	ctgggattac	aggctctgagc	cgcgcgccc	ggcctagaat	81660
atttgtttcg	attattctaa	atctggtgac	atttcttttg	tttttaagtt	aaatcttcag	81720
taaaaagaat	aaatgccacc	tagaggacag	aaaaattttt	acagtagatt	atcacagacc	81780

tcattgactca	ttacttttgg	gtataaaatg	gcctttgtat	ggtgtcagca	cctgggaatg	81840
tctcaagggg	tggttcattga	ccttctggac	tatctggaaa	tattttgata	tttattggct	81900
gggcgcggtg	gctcatgcct	gtaattccag	cacctcgga	ggccaagtta	ggtggatcac	81960
ttgaggtcag	gagtttgaga	tcagcctggc	caactaagtt	agttgactat	ttgtgaaatt	82020
tccctctcca	aataggaaag	ataggattgg	aaacattatt	cagaaggaag	aagttttaaa	82080
gaagggcagg	cttaaagact	atctaataaa	tttaccata	tggtaaaacc	ccatctttac	82140
taaaaataca	aaaattagcc	aggcgtggtg	gtgggcgcct	gtaatcccag	ctacctggga	82200
tggttgaggga	tgagaatcac	ttgaaccag	gaggcaaaag	tcacagttag	ccgagattgc	82260
accactgcac	tccagactgg	gctatagagc	aaaactctgt	ctcaaaaaaa	aaaaggaaaa	82320
aaaaaaaaga	aattttgata	tttatgtgag	aatgactttt	cacggtgttc	ttaatagcgc	82380
aagttttgtt	taggagagca	cattccaact	tacttgcttc	tataaatata	ccgtgtaatc	82440
tagggcttaa	agaatatgta	caatgtcttt	ttctctcccc	tagtcttccc	cttttctcac	82500
tagttatatc	cttgactgaa	gagctatttc	attctcaagt	cttaggaatg	cagggtgaag	82560
caaacacagat	gaatttttaa	agcacttaac	caggctgtat	gaaatcacag	tctgttgtct	82620
aaaattgtaa	gggacgggta	gtagtactct	cccctttctc	aaaccaaata	tttggttgtt	82680
ttaaggaagg	gaaattaaat	tcctgtgtgc	tagattttca	acataaaatt	taaaaaactg	82740
gaataatggt	gcatttttatg	tgtataacag	tataatgctt	tgtttttaggt	gggagaagga	82800
gagaaattgg	tgagggtctc	ttttgctgtg	gctcgagaac	ttcaaccttc	tataattttt	82860
ataggtaaga	acatattttc	caactaagtt	attgactatt	tgtgaaattt	ccctctccaa	82920
ataggaaaga	tacgattgga	aacattattc	agaaggaaga	agtttttaaag	aagggaagc	82980
ttaaagacta	tctaataaat	ttagtaggac	ccactatatt	aataagtagt	aaactagatt	83040
aatctcagat	gactcacata	gcttggtctt	taattaaagt	cttatacttg	tatttcctct	83100
agatgaagtt	gatagccttt	tgtgtgaaag	aagagaaggg	gagcacgatg	ctagtagacg	83160
cctaaaaact	gaattttctaa	tagaatttga	tggtgtaagt	ggtgattatg	atatttttaa	83220
tgtggcagca	tttttagtata	ttttcctatt	aaatggccaa	ggttaaaaaa	acaaatatct	83280
ttatatattgt	tattactttt	ctaaatgaat	tgaaaaaaga	ttttttgctt	gtaggtacag	83340
tctgctggag	atgacagagt	acttgtaagt	ggtgcaacta	ataggccaca	agagcttgat	83400
gaggctgttc	tcaggtaggg	agatttatat	ggaaatacat	gcatttatta	cagacaatat	83460
ttactcatgt	gtccatctta	catattattt	ccttactctc	agttttaaga	ctaaattcac	83520
tattttcttc	cagtactatc	tctagcctct	tgttaccaac	tacataagga	ttttgaggtc	83580

aacagcttga	tatcaggaga	attagtctag	taaaggttta	gttacggttt	taggcaaatt	83640
ggtcagttct	tctgggttgc	attaaattat	ctttaaactt	gaaaattgat	tcttaccgtc	83700
ttttattggg	cttcatgaga	aaataatcaa	atcagggttc	atggaaaatg	tgttttctgt	83760
tcttgtgtgt	ttgagcagtt	tattcccttc	atactggaag	gacggtttgg	gtcagcataa	83820
aattcttggg	catcctttcc	caaggaatth	acaatattgc	tccattgttt	ctagattggc	83880
tgtagtagtg	taaaagtctg	aatccaaatt	gattttttca	aatttgtaaa	ataacttgat	83940
ttttttctct	ctttactgga	tgggttacgt	gggtatatgt	caccaggtta	gtgagcatag	84000
ggtgcagtag	gtagtttttc	aaccacacac	tcactgcttt	cttcccccat	gtagtagtcc	84060
gcagtgtgta	ttgttcccat	gtttatgtcc	atgtgtactc	aatttttagc	tccccactta	84120
tgagaacatg	agctatthgg	tttactgttc	ccacattaat	tggcgtagta	ttaaggcctc	84180
cacctccatc	catgttgctg	caaaggacat	tatttcattc	ttttttatgg	ctgcgtagta	84240
catagtattc	catggtgtgt	acgtaccaca	ttttctttat	ccagtccacc	attgatggac	84300
acttagattg	attctatgtc	tgctgtcatg	aatagtacag	caataaacat	gaaatgcatg	84360
tctttttggg	ataatgatct	atthtcctth	gggtatatac	ccagtaatgg	gatttcaggg	84420
tcaaaggtag	atthgtthta	tgthttttga	gaaatcttca	aactgctthc	tacaatgact	84480
gaactthtaca	ttcctaccag	cagtacatac	gctccacaac	ctcaccaaca	tctgttattt	84540
ttttacttht	tcccgccaat	ctgtagacaa	tatgggattt	tttttgccat	tttattaata	84600
gccattctga	ctggtgtgag	atatctcatt	gtgattttga	tttgcattht	tctgatgggt	84660
agtgatgatg	agcattthtt	catgtthgtt	gtcccctcgt	atgtctthtg	agaagtgtct	84720
attcattthc	ttthtaccca	ttthttaata	gggttggtgt	ttthtagctt	gttcaattgt	84780
ttgaattcct	tatagattct	ggatatcaga	ccattgttag	atgcagttct	gtagattgtc	84840
tgthtactct	gttgatgggt	ctthttgctg	tgcagaagct	ctthagttht	attaggtccc	84900
acttgtcaat	ttthatttht	gttgcaattg	tgthttggga	cttagccaaa	atttctthct	84960
gaggctgatg	tcaagaagtg	tatttcctag	ttthcttcta	gcattthttat	actthtgagat	85020
cttacattta	aatctthaat	ccaccttgag	ttaattthtg	tatatggtga	aaagtaaggg	85080
tccagthtca	ttthtctaca	tatggctagc	cagtaatccc	catgccattt	attgaatagg	85140
gagtccttht	cctcattgct	tgthtttgct	agtcttatca	aagatcagat	ggttatagggt	85200
gtgtggctth	atthctcaat	tttctgtcct	gttcctthtg	tttgtgtgtc	tgthttttata	85260
ccagaattat	gctgtthggg	ctgctgtgta	gttcggthta	aagtcaggta	acgtgatact	85320
tacagctthg	ttctthttgc	ttaggattgc	tttgtctatt	tggactctth	ttthgctthc	85380

atatgaattt	tagaatattt	ttttctaatt	ttgtgaaaaa	caacattgat	agtttgatag	85440
aaataccatt	gaatctgtaa	attgcttttg	gcagtatagt	cattttacga	ccagataaat	85500
caggcaagag	aaggaaacaa	aaggcatcca	aataggaaaa	gaagtcatca	tactctcact	85560
cttcactggc	aatatgattc	tatactttga	aaaccctaaa	aattccgtca	aaaggctact	85620
agaactgata	aatgatttta	ccccaaagatt	caggatacaa	aatcagtgtg	caaaaaatta	85680
gtaccatttc	tatacgctaa	taatgtccag	gctaagagtc	aaatcaagaa	cacagtccca	85740
tttacaatag	ccacaaagaa	aatgaaataa	tggctgggca	cagtggctca	tgctgtaat	85800
cccagcactt	tgggaggctg	aggcagggtg	atcacctgag	gtcaggagtt	cgagaccagc	85860
ctgaccaaca	tggtgaaacc	ccgtctctac	taaaaataca	aaaaattagc	caggcgtggt	85920
ggtgggcgcc	tgtaatccca	tctgctcggg	aggctgaggc	aggagaattg	cttgaacca	85980
ggaggcatag	gttgacgtga	gccgagatcg	tgccactgca	ctgcagcctg	ggtgtgaaag	86040
agcaagaatc	cgtttccaaa	attaaaaaaa	aaaaaagaaa	aaaaaaaaga	caatggagta	86100
cctaggaata	cggctgatga	aggaggtgaa	agaatctgta	tgaggagaag	tgtaaagcac	86160
tgctaaaata	aatcagaggt	gacacaaata	aatggaaaaa	cactccatgc	tcatggattg	86220
gaagaatcaa	tatcgttaaa	atgggaaatt	tatcttaatt	tttttccta	gttcattttt	86280
ctgttttctt	ttgcacataa	attatgtggt	atgttgcgtc	cttttgtttt	ccctatctgt	86340
aatttttctt	ttcatccttt	taaactctta	ggtctttttg	ttttgttctg	ttttcccaag	86400
cctgttgcc	ttattttctta	ctgggttagt	atctaattct	ctttctgcta	tatcttttgt	86460
aactcttctt	tccttggtgg	ttttttattc	attactttcc	tgacatctgt	tattctactt	86520
ttcatcttct	gttttatttc	tttgagcatc	tgtatgttat	ttccataagc	tcttacattt	86580
cactgggctc	ttgtttcata	gaaaacaatt	ttttggcaag	gacagaccac	ttgaggccag	86640
gagtttgaga	ccagcctggc	caacatagta	caaccctgtc	tctgctaaaa	atacaaaaaa	86700
ttagctgggt	gtggtcacac	acacctgtaa	tcccagctac	tcaggaggct	gataatcgct	86760
tgaacccagg	aggcagaggt	tgacgtgagc	caagacggcg	gcactgtact	ctagcctggg	86820
ggacagagca	aggctctgtc	tcaaaaaaaaa	aaagaaaaaa	ttttttggtc	aggcgtcatg	86880
gtgggtcaca	cctgtaatcc	cagcactttg	ggaggccaga	tacttgagg	ccgggaattc	86940
cagaccagcc	tcgccaacgt	ggcaaaaccc	catatctata	aaaaaaactt	aaggataaaa	87000
aaaattaaat	ttttaaaaaa	ttaaattatt	ttatttcgtg	gaaagatact	tggtcataat	87060
tttagtttgc	tctatggcaa	aatcttctca	tgcatgttct	tcattttatc	tagtttatgt	87120
ttattgtgtt	ttgtttgctc	ttttttgggg	gcagggtttg	ttgttggtgt	tttttttttt	87180

ttttttttaa	atgagccctg	tcctgggttt	tccttttttg	cttattat	ttctttgaat	87240
aaaaggagtt	ttcttgga	acctagtttt	ggaaggatat	tgtaggggaa	gggataggta	87300
tcttgagtag	tccatgtata	cactaaacta	gcatgcagcc	tccttcactt	acagtgaagc	87360
cctaccta	aatgaagt	atgtgtatgt	aaataagaga	tagggttgat	ttattcttct	87420
actaaaaa	attggtcact	gcaagacatt	gtcttccgc	cagactgtta	gatgccttct	87480
tataaagata	attcgtctta	gagtttcttt	gttcattctc	accttccttg	ccattcctat	87540
gctatcaatt	agggtttata	tggtctgcag	agccaggctt	catatgcttg	tttaa	87600
gagggactgg	ccttattacc	ctttagggtg	tgctatctct	ttttgggagg	aaaactgtac	87660
tctgacagtc	atgtccatga	tctttagcgg	cagatccttt	gtgtgtcttc	ctttgtcagt	87720
gtaattttca	ctgccctagg	cagttcacct	ttatgtat	tagttccata	aatgttacct	87780
agtttggtga	tgaagtgtat	gtttctgttt	ctattatcct	tgctgatttt	agtttttttc	87840
agattcaaga	agacagatct	acttatatca	gatattttgt	ttgttttaaa	gcattactgc	87900
tattaaaaac	atatacat	ataaatgatt	aattctaaca	ataccgaatt	aaggatttta	87960
aacattaact	agagttttaa	aagagtaaca	aaaatatttt	tacattgata	actacaaaa	88020
tgcatccag	tgccctgaat	attatatttt	aaaagctttt	cctgtcattt	gctgtttcag	88080
ctttaaattc	aaaattatat	ttctaaaagt	gctggatttt	tttttttagg	cgtttcatca	88140
aacgggtata	tgtgtcttta	ccaaatgagg	aggatgtat	ctgtgtttga	at	88200
ttttagagca	gaaacaagaa	ctaccatctt	gacaatatta	agtcttccaa	tccatggtac	88260
agctactttg	gaaaacagtt	tagtggtttc	ttaaaaagat	gtacataagc	ttactgtcag	88320
agccagcaat	ttcactccta	agaatttacc	ccagagaaac	aaaaatgtat	gtccacacaa	88380
agacttgtag	aagaattttt	atagcagcaa	tattaataat	agccagaact	acaaatgatc	88440
taaaatgttt	at	ataaacaat	gtggtatatt	catttaatgg	atactgttat	88500
ccattataca	tgctgcatac	tcacttcata	tattaagttt	tatgaatgaa	actccaaaac	88560
attataagtg	aaagaagcaa	gatgatatat	gtcgcatatc	ataggattct	gtttatatgc	88620
agaaaacgca	aatttatttt	ttatttattt	at	gatggagtct	tgctctgtcg	88680
ctcaggctgg	agtgcagtg	cgtgatctcg	gctcactgca	accttcacct	cctgggttca	88740
agtgattctc	ctgcctcagc	ctcctgagta	gctgggactg	caggcatgtg	ccaccaggcc	88800
tggtctaattt	tttgtgtttt	tagtagagac	gggtttcact	gtgttagcca	ggatggtgtg	88860
atccgccctc	ctcagcctcc	caaagtgtctg	ggattacagg	catgagtcac	cg	88920
cagaaatcga	aaatttctag	aaacgtaaag	cagatcagtg	ggttgtcagg	gtgagagttg	88980

ggaatgagga	ttaactgcaa	acaggtatga	gtgaactaaa	aagtgttcta	aaactggatt	89040
gtggtgatga	ttgcacatct	ataaatgtat	taaaactcat	tgaattatat	acttacaatg	89100
ggtgaaattc	atagtagatg	aattcatacc	tctataaaac	tggttttttg	cagcaaaata	89160
tttattgcta	tggtttttatt	caagtgtggt	gaattattgc	tatgttttta	tccaggtgtg	89220
gtgtctcacg	cctataatcc	caacactttg	ggaggctgag	gtgggaggat	cgcttgagcc	89280
aggaattcaa	gaccagcctg	ggcaacatag	ggagaccctg	tctctacaaa	aactttttta	89340
gaaattaacc	aggcatggtg	gcacgtgcct	ttggtcccag	ttactcggga	ggctgagatg	89400
ggaggattgc	ttgaaccag	gatatcgagg	ctatagttag	ctatgattgt	gccctgcact	89460
ccagcctggg	taacagcaca	agaccctgtc	tcaatataaa	aaaagaaaag	aatcattaat	89520
tctgaaatta	gactgaatga	tcatttttta	atatttttca	gacaagacta	cttttgctta	89580
aaaatctggt	atgtaaaca	ggaagtccat	tgacccaaaa	agaactagca	caacttgcta	89640
ggtgagtaat	ttggatttg	tttatcttac	agcttttatt	tattttttgt	aaataattct	89700
ttttttagtt	atttaaagta	atcttaagta	tgaaatgagt	aattcattga	tcagaagact	89760
ttctctcatc	ctctacctcc	tttgcccttt	tcttacctct	tggtcttata	tatatatgtg	89820
gaatctaaat	ttcactggct	atgtcctttg	caagacgtga	tctaattgatg	atagtatat	89880
ctttttataa	atttattaaa	acctgtaagt	ggattataag	taatttaaac	atttacacct	89940
taggcaacat	ggtaaaacc	catctctaca	aaaaataaaa	tgagtcagga	atggtggcac	90000
atacctgtag	tcccagctac	ttgggaggtt	gatgtgggag	gatcgcttca	gcccgaggag	90060
tgagggttga	gcaacagagt	gagaccctgt	ctcaaaaaaa	attttaaaca	tttcatattg	90120
cacactaaca	cagatatatt	tatataaact	tggtcttggc	atatgctgaa	tacttttcag	90180
ctcttttaat	aatttttaatt	atttgataat	tttaataatt	ttaatttcag	ctatttgaat	90240
aatcttaaac	tttatgaaca	gtgtgaagta	gaaactgagt	ataggagatg	cgctcatgaa	90300
ttttattccc	tgaaaaatgt	cataaagtaa	tataaagtat	tcaatgagta	tgtgagtatt	90360
gaataatttt	aacaagaact	ctgagttccc	agatacctat	gacaaaagaa	caaattactt	90420
ggaagtctct	tctactccag	acctttctgt	ctatcccctt	atttaaaaaa	atgtacattg	90480
attgtccctt	atcccatcaa	gagagtgttc	tttgccaaga	taaacgctaa	ttagagaagt	90540
aattgtgttt	atgtgtcagt	tggccaccag	tagttttag	atgttacctt	tccaggtgac	90600
atgtttacaa	tttaatgaag	cctctggaga	tagtatgcct	taatctaaag	gaatggttct	90660
aatgattaaa	atttgaggca	ttactacatt	tgttttcagc	aaatgggcct	ttctattccc	90720
ttaaaatcaa	accctcataa	ctatgggttt	gctgacaaa	ggaaactagt	ttttactctt	90780

attatatgga	gatttttaaat	ctgcagagat	gaaaataatt	tgctgtttca	gctgggtgcg	90840
gtggctcatg	cctgtaatct	cactactttg	ggaggctgtg	gcgggtggat	cacctgaggt	90900
caggagttcg	agaccagcct	gaccaacatg	gcgaaacgcc	atctatacta	aatacaaaaa	90960
aaaaattagc	caggtgcggt	ggcgcatgcc	tataatccca	gctacctgag	aggctgaggt	91020
aggagaatcg	ctccaggagg	tggagatcac	agtgagctga	gatcatgcca	ttgcactcca	91080
gcctgggcaa	caagagcaaa	actccatctc	aaaaaaaaagc	gggaggggaa	ataatttgct	91140
gtttcttcct	tcccttcctc	agaatgactg	atggatactc	aggaagtgac	ctaacagctt	91200
tggcaaaaga	tgcagcactg	ggtcctatcc	gaggtaggta	tacaagagct	taaaacattt	91260
agaactattt	attataccac	cttagaagtt	taagaagtcc	aaaaaaatct	accaagagat	91320
tttttttttc	ttttggagac	agggctctac	tctgttgccc	aggctgtagt	gcagtggcac	91380
aatcatagtt	cactgcagcc	tcgaaccccc	caggctcagg	tgattctccc	accagcctt	91440
cccagtagct	gggactacag	gcatgcacca	tcatgccag	ataattattt	tattttattt	91500
cttgtagaga	cagggtttca	ccatgttgcc	caagctggtc	ttgaactccc	gggctcaagt	91560
gagcagcctg	ccttggcctc	ccaaagtgtc	gggattacag	gcatgagcca	ccatgccag	91620
ccttaaaacc	attcttagct	cacagatcat	acaaaaacag	tctatgagcc	agtagactgt	91680
gaccctgtt	ctaggcaaga	ggtttcattt	cttgatagta	agtaagcagc	agaaaactca	91740
aaaagaaaac	tttaaacaaa	aaagaataat	gtaaggaaaa	tgttcttata	tgttttctgg	91800
ctcttttcag	tattctgaca	tcatttgaaa	ttaatgtgtc	taaaggaaat	tctataagag	91860
aagttttcac	ttttttaatt	atatcttttg	agggattgca	gcaaaatatt	tattgctatg	91920
tttttattca	ggttacttgg	ttgattctga	aataccacga	ccaattcttt	ttaggttagg	91980
ttaggataaa	ctcacacaga	tattttcagc	tacatttcca	cagccagcat	cggtagtgga	92040
tggttgatgc	tctcaggtat	gcactcagat	tattagctat	gatgataaag	taatcataat	92100
agctccatcc	ctttctctca	ttcaagaaag	tacaattgaa	tgtagttgta	cgagaaagat	92160
gttatcatag	tgtctactgt	aatttgcat	aaaagcctac	attatacaaa	cctttttgtt	92220
ttactagtta	aaaattacca	gtaacacctc	ccgtagtga	ataggggtgct	gattaagaac	92280
tgctactttg	caaaataaaa	aatgtaatct	tgagagtata	ttttgaaact	ctggatgaac	92340
aaattttctt	gctgctggag	cttaaactct	gttcagttag	tggatttaga	aacagtagct	92400
aatgcataaa	tgagttgtat	tttcttagaa	cgttttggct	attctaaggt	agacaaaatt	92460
tgggtttaaa	aaaagattcg	catccattaa	agtacagaaa	atggccgggt	gcagtggctc	92520
acacctataa	tcccagcact	ttgggaggcc	aagggtggag	gatcgtgtgg	gcctaatagt	92580

ttgagactag	actgggcaac	atagggagat	cctgtctcta	caaaaaata	taataaatta	92640
accaggcccg	ttggcatgag	cctgtaatcc	cagctactcg	ggagtctgaa	gtaggaggat	92700
cacttgagcc	tggaaggtgc	aggcagcagt	gagccatggt	catgccactc	actgcactcc	92760
agcctgggtg	ggtgacatag	tgaggtgctg	tcttaaaaag	taaataaata	ggccaggcac	92820
ggtggctcac	gcctgtaatc	ccagcacttt	gggaggccaa	gatgggtgga	tcatgaggtc	92880
aggagttcga	gaccagcttg	gccaatatgg	tgaaaccgca	tctctactga	aaatacaaaa	92940
attagccagg	cgtggttgtg	cgcacctgta	atcccagcta	ctcaggaggc	tgaggcagga	93000
gtatcgcttg	aacctgggag	gtgaggttgc	agtaagccga	gcactgcagc	ctgggcaaca	93060
gagcaagact	ctgtctcaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaagtaa	atacataaag	93120
tgcagagtaa	aaacaaaaaa	agactaatgc	attttgtaaa	gaacaagttg	cattctttta	93180
agttcccttt	caaatttggtg	aactattggt	tttgggcagt	atgcaagaaa	ttgaacactt	93240
tccagttatt	ccaaagaagg	aatattctct	tctcagcatt	tataaattgt	atttgctctc	93300
aaagttaaca	tgtgtctctt	ttttttaata	taatgatttg	tactgaatag	atacatgtag	93360
atcattgtac	ttggttttgc	ccttcaacaa	tttcaactgc	aaaatgtatg	tatttttaag	93420
tgcctgactt	ttatgtttta	cagaactaaa	accagaacag	gtgaagaata	tgtctgccag	93480
tgaggtatag	tattttacaa	tgatattttc	tttgtcttct	atattgtaag	acatatataa	93540
gacatacata	tatgaatgtg	tgtgtgtgtg	gttttttttt	tttttttttt	ttttttgaga	93600
cagtctggca	tgatatcggc	tcactgcaac	ctccacctcc	tggattcaag	ctattctcct	93660
gcctcagcct	ccaagtagc	tacaggcatg	taccaccaag	cccagctaata	ttttgtattt	93720
ttagtagaga	gggcatttca	ccatgtttggc	cagtctggcc	tcagactcct	gacctcaagt	93780
gatccacca	actcggcctc	ccaaagtgct	gggattacag	gcttgagccg	ctgcatctgg	93840
ctgttttggt	tgtttgtttg	tttgtttggt	tgtttgagac	ggagtctcac	tctgtcgccc	93900
agactggagt	gcagtgggtgc	gatctcagct	cactgcaagc	tctgcctccc	gggttcacgc	93960
cattctcctg	cctcagcctc	cctagtagct	gggactacag	gcgcccacca	ccacgcctgg	94020
ctaatttttt	gtacttttag	tagagacggg	gtttcaccgt	gttagccagg	atggctctctg	94080
tctcctgatc	tcgtgatctg	cctgcctcag	cctcccaaag	tgctgggatt	acaggcgtga	94140
gccactgcgc	ccggcctttt	tttttttttt	tttttgagac	acagggtcct	gctctgttct	94200
caggctggag	tgcagtgggtg	caatcgtggc	tgactgtagc	cttgacctct	tggctctcaag	94260
cagtccctct	gtctcaaact	tttttagtagc	tgggactaca	ggcaagtgcc	acaacgcccc	94320
actaattttt	taattttctg	tagagatggg	gtctctccct	gttgcccagg	gtgggtctcag	94380

actcttggcc	tcaagtaatc	ctcctgcctc	agtctctcag	agtgcttggg	actaattttt	94440
tttaatggat	ttttctagtt	gattagaagc	tagaaaaaat	taactttgct	tcattttcaaa	94500
atatggaaac	ctatgaaata	gtcaaaaattt	tttttggttaa	tagacaaaac	atatatacat	94560
aaagtcattt	tctatcaaata	catattacat	aaataccaaa	cagaagaaaa	ttacaggccg	94620
ggtgcggtgg	ctcacgcctg	taatcccagc	actttgggag	gccgaggtgg	gcagatcact	94680
tgaggtcagg	agttcaagac	cagcctggcc	aacatggtga	aaccctgtct	ctatcaaaaa	94740
tacaaaaaaa	ttaggccggg	ggcgggtggct	cacacctgta	atcccagcac	tttgggaggc	94800
cgaggcgggc	ggatcacgag	gtcaggagat	cgagaccatg	ctggctaaca	tggtgaaacc	94860
ctgtctccac	taaaaaatac	aaaaaaatta	ccggggcgctg	gtggcgggca	cctgtagtcc	94920
cagctacttg	ggaggctgag	gcaggagaat	ggcgtgaacc	tgggaagtgg	agcttgccagt	94980
gagccgagat	cgtgccactg	cactccagcc	tgggcgacag	agcaagactc	catctcaaaa	95040
aaaaaaaaaa	atagctgggc	atggtggcac	gtgcctgtgg	tcctggctac	ttgggaagct	95100
gaggcaggag	aatcacttga	acttgggagg	cagaggttgc	agtgagccaa	gattgccccca	95160
ttgcactcta	gcctgggcga	cagagtgaga	cccagtctca	aaaaaaaaaa	aattacatga	95220
aacttggtgg	ggtttataca	ataaaaaataa	cctataacta	tacgttaatc	ctttaataat	95280
gttctatttt	gtttttcaat	aacctttatt	tttttaatta	atttcttttt	tttgagaagg	95340
actctggctc	tgtcacccag	gctgcagtgc	agtggcacca	tcttggctca	ctgcagcctc	95400
caacctccag	gttcaagcga	ttctcctgcc	tcagcctcct	gcgtagctga	gattacaggc	95460
atgggccacc	acgcccagct	aatttttttt	ttttttgtat	ttttagtaga	gatgggggttt	95520
cactatgttg	gccagggtgg	tctcgaactc	ctgacctcag	gcgatcgcca	gcctcggcct	95580
ccctaagtgc	aaggattaca	ggcgtgagcc	accatgcctg	gccaatatat	taccgtttct	95640
taatatggaa	gacctcctta	tgagatgctg	aagcatttca	tttgaaaaaa	gttgcatgta	95700
aaatataaaa	ttgggtatgc	agaatggtca	taactctgtg	agcaaaaattt	tgaaatcagg	95760
cttgaattac	agtcagcctt	ttctatccgt	ggattctgca	ttcatggatt	caaccaactg	95820
cagatcaaaa	atacttgggg	gaaaaaagca	tctgtgttga	acatgtacag	acttttattc	95880
ttgttattat	tccctaaaca	atatagtata	aaaactattt	acattgcatt	tacattgtat	95940
taggtgttat	aagtaatcta	gagatgattt	aaagtataca	ggaggattgt	gtaggttgta	96000
tgcaaatact	gcaccatttt	acataaggga	ctttaagcgt	ttgcggattt	tgctatctat	96060
gggggaacca	atttccgaga	gatactgagg	gacagctgta	tatttgtaac	ttatttttta	96120
tttccctaata	tgcagcagct	gttgagggga	cagtgaactg	ttaacacaga	taacaagtgt	96180

attcaagtac	atTTTTggagt	tTgTtTgTtT	tTtTtTtTtTca	ggcTtTtTatt	cgCctgtCgG	96240
atgaggcacc	ataCttgaat	tTtTtTtTtTtT	aagaaagCtt	ttagTtTtTct	tTcTtTtCtTt	96300
tTtTtTtTtTtT	gtgagatgga	gtCttgCtCt	gTtTtCtaggc	tggagtgGag	tgcagtgGtg	96360
tgatCtcagc	tCactgcaac	CtCcgCctac	tgggtTcaag	tgattCtCct	gcCtcagCct	96420
cccgagtagc	tgggactaca	ggtgcacgC	accacgCcca	gCtaattTtTt	tTtTtTtTtTt	96480
tTtTtTtTtTg	tattTgtagt	agagacagag	tTtCaccatg	tTggccagga	tggTctCgat	96540
CtCctgacCt	tgggatCcgC	CcaccTtggC	CtCccaaagt	gCtagggatt	acaggCcatga	96600
gccaccagC	CgggCtgaaa	gCtTtTtagTt	tTctaaCtTa	tTtaattTtaa	tTtaattTtaa	96660
tTtattTtTta	tTtTattTtTt	tgggacagtg	tCtCacttgG	tTgCccaggc	tggagtgCag	96720
tggTacaatC	atggCtCact	gcacCctCta	gCtCctgggC	tCaggCaatC	CtCctgCctC	96780
agCctTtTga	gtTgCtggga	CtacgggCcat	gtaccaccac	actCagCtaa	attTtTtaatt	96840
tTtTgtagag	atggggTctC	actatgTtgc	CtaggCtggT	CtCagattC	tCaagCattC	96900
CtCccactTg	CacCtCccaa	aatgCtggga	tTacaggtgt	gacacCgtgc	CagactTgaa	96960
attTtTaatC	CcacacCtaa	aaatataatt	tTatCcacCa	tTtTtTaaaa	gtcataacat	97020
tattTattaa	aaattTtagat	ggTaaaacta	aaaattTaaag	CtTaataaaa	Ctactgagtg	97080
attTacaagg	aagaatatta	CtggTccCtT	tTgtgagCcat	CccattTaat	atatataTtC	97140
aggTtatatt	gtTacaatat	tTggTtCtac	tgtataCtTt	tTtTtTtgag	actgagTtTc	97200
gCtgtTgtTg	CccaggCtgg	agTgCaatgg	tGcaatCtCg	gCtCactgCa	acCtCcgCct	97260
CctgggTtCa	agTgattCtC	CtgcCtCagC	CtCctgagTa	gCtgggatta	CaggCcatgCg	97320
CcactatgC	tggCtaattT	tgtgtTtTtTg	tagagacggg	gtTtCactat	gtTggTcagg	97380
CtgatCtGga	actCccaacc	tCaggagatC	CgCccacCtC	ggCctCccaa	agTgCtggga	97440
tTgcaggtgt	gagccaccat	gcCtggCtCt	actatataCt	tTcattCagt	tgtTtCtTtT	97500
taatCtagtg	gtTtTggTat	taataattTg	ataatgacCt	ttagCtgtTa	tTgCttactT	97560
atgagTtaat	attTataaag	CactTaaagt	agCtggCaCg	tagTaaacac	tatgtTaaaga	97620
tccattTaat	aactTtTaaaa	aatataaaaC	tGatagTggC	attTtattat	agagattTaaG	97680
gTaatccatt	CtCtCattTc	CattTatggg	atgagacgTa	aacacaagTa	gtTtGtCtC	97740
TaaaactgTa	tactagaatt	tTatataCcc	attattTgat	gCaactTtaa	tacCaaagtG	97800
tattgtCagt	taccggtgaa	tatatataaa	tTtaggtTaaG	gaaaacCcaa	CttggTcGca	97860
CacagTggCt	CacgcCtGta	atCctaccac	tTtgggaggc	CgaggCagTt	ggatCacCtG	97920
aggTctTgag	tTcGagacca	gcCtggCcaa	CgtggCgaaa	CccCgtCtC	actaaaaata	97980

caaaaattag	ccgggcgtgg	tggcacatgc	ctgtaatccc	agctgctcgg	gaggctgagg	98040
aaggagaatt	gcttgaacct	ggggaggtgg	aggttgacgt	gagctgagat	cgtaccactg	98100
cactccagtc	tgtgcgacag	gagcgagact	ccatctccaa	aaaaaaaaaa	aagaaaaacc	98160
caacttatct	tttacagttt	ataatagtag	aagttcaaat	aattggtttg	gaatttctgt	98220
atttttaagg	ttagtactaa	aattgttggt	tataaattgg	ggtacaatat	actttgtttt	98280
ttaagtacct	tgtgtatcta	atttaacttt	aagtccttta	ttattttggt	ttgataagac	98340
aactttttcta	cttattttccc	cccttaactg	aaccagctac	catctgcctt	tttctgtttg	98400
tatacattag	tctcttacgt	taaaatatca	tataagtttc	atatatatac	acatatcaaa	98460
ctatagactt	aaagtacaat	tacatcaaca	tctttttaaaa	ccttaatttc	tggccaggcg	98520
tggttgttca	tgcctataat	cccagcactt	tgggaggctc	aggccattgg	atcccttgag	98580
ctcaggaatt	cgagaccagc	ctggataaca	tggtgaaatc	ctcatctcta	caaaaattac	98640
aaaaagttag	ctgggcgtgg	tggcgcacac	ctatagtcct	agctacttgg	gaggatctct	98700
taagcctggg	aggcagaggt	tgcagtgagc	cgagatcatg	ccactgcact	ccagcctggg	98760
taacagtgag	accctgtctc	aaaacaaaaa	tattaatttc	tataacaaat	aaattttatt	98820
tagtaatatc	atgtttaata	cccatgttac	attcattatt	cttctacact	ttggccttta	98880
ggaagaggca	gtgtgtgtgt	tacagttaaa	taggaaacaa	gtagggttca	tatagtgcct	98940
atgggggtttt	ctttgggtggg	gggttgctgg	gtttttgggg	tttttttttt	tgagacaggg	99000
tatcactctg	tctcccaggc	tgaatacat	tggcatgctc	atggctcact	gcagcctcaa	99060
cctcccagac	tcaagtgatc	ctcccacctc	agcttcoccta	gtagctgaga	ctacaggcgt	99120
gcgccaccat	gcctagctaa	attttgtaca	ttttgtagag	atgaggcttt	gccatgttgc	99180
ccaggctggt	ctctaactcc	tgggctcaag	tgatccactt	gcttcagcct	cccaaagttc	99240
tgggattaac	aggtgcgagc	cactgcagcc	agcctatatt	ttaactatat	gtttttttct	99300
ttttggctaa	aatttttccag	attagtttac	aagttacaag	tgtaggatgat	atctcatgga	99360
gatatcaaag	atgatatgaa	gttagattgg	gtttttaaga	gtagttttta	aaatacggat	99420
aaataccagt	tgttggagtt	ttgtttaaag	ttctttaact	tcttattttt	gggccaaaat	99480
acaggtatac	gctagaaatg	atttttaaca	caggtcattt	atgccaaact	gcattttgcc	99540
ttaattttttt	tttttttttt	tttttttgag	agggagtctc	actcattgcc	caagctggag	99600
tgcaatggca	cgatctcggc	tcacttacaa	cctccacctc	ccaggttcaa	gcgattctcc	99660
tgcctcagcc	tcctgagtag	ctgggattac	aggcacctgc	ctggctaatt	tttgtatttt	99720
tagtggagac	gggggttcac	catgttggcc	aggctggtct	cgaactgctg	acctcagggtg	99780

atccacccgc	cttggcctcc	caaagtgcgg	ggattacagg	tgtgagccac	cgtgcttggc	99840
caaaatTTTT	attaatTTTT	ctattgcctg	gactctgtga	acctatccat	tttgcctttt	99900
aaaaatactt	aggtgtaaat	atagatatct	attaactcag	cattgtttta	atctatatatt	99960
ccaaaggcaa	tttaaaagat	cagaaaataa	gaccaaatta	atataaaaaat	gcataacttta	100020
ggctgggcaa	agtggctcac	gtctgtaatc	ccagcacttt	gggaggccaa	gggtgggcgga	100080
tcacctgagg	ttaggagttc	aggaccaacc	tggccatcat	ggcgaaaccc	tgtctctact	100140
aaaaatacaa	aaattagctg	ggcatggagg	catgtgcctg	taatcccagc	tactcgggag	100200
gctgagacag	aagaattgct	tgaaccttgg	agggggagggt	tgcatatctg	agtgggtgaaa	100260
ttgtgattct	ttttttctct	ttgtctgtat	ttttgaactt	ttctataaat	gattgtgttt	100320
tgtttttata	ttggaaaaat	attatgcttt	caaagtgtta	tacctatgaa	actaaacaca	100380
agtaataaat	atattagtat	agcatttatt	aaggtttctt	gtgtagcaga	tcaacataga	100440
aaatatattt	aatggctga	cataattttc	taagaataca	tacacgtata	ttttttataa	100500
cattaagaaa	cagcagcatc	attactttta	tccatcattt	cgttaaccac	catatacctg	100560
ttgatcattt	gtattgtcat	gtgcttttta	aaaatctaga	tgagaaatat	tcgattatct	100620
gacttcactg	aatccttgaa	aaaaataaaa	cgcagcgtca	gccctcaaac	tttagaagcg	100680
tacatacgtt	ggaacaagga	ctttggagat	accactgttt	aaggaaatac	ctttgtaaac	100740
ctgcagaaca	ttttacttaa	aagaggaaac	acaagatctt	caatgaacgt	catcggctac	100800
agaaacagcc	taagtttaca	ggacttttta	gagtcttaca	tatttggtgca	ccaaacttga	100860
agatgaacca	gaaaacagac	ttaaacaaaa	tatacaatgc	aaatgtaatt	ttttgttggt	100920
taaggccttg	ccttgatggg	cacagttatc	ccaatggaca	ctaagttaga	gcacaacaaa	100980
acctgattct	ggctcttctt	accaatataa	tcataatgta	aataataatt	tgtatattgt	101040
gttgacagatg	aaagtattcc	aggaacagtg	aatggtagaa	gacacaagaa	catttgtttg	101100
tttgccttct	gatgtttttt	cttaaaatag	taatttctcc	tacttttctt	ttctactggt	101160
gtcttaacta	caggtgattg	gaatgccaaa	cactcttaag	tttattttct	tttttcgttt	101220
tataaattca	gtgtgccaaa	tgaaactttt	ttcctaagta	actgtaatag	gaaaaagttt	101280
atTTTgagag	tttcttcttc	ataaatctac	agacattaaa	caattgttgt	gttcttttta	101340
ccttttattt	ttctattacc	ttgctaccaa	acagtttaga	tagcaatata	atagcaaaaa	101400
agcaaatatg	gtaaaaataga	gaaggtttga	aggtttgagt	tactctgtca	tataacatgt	101460
agatcagtct	tcatgtgacc	tgcagtattt	ttttttctaa	tgtatttgtc	agaaatctgt	101520
tgtagactgt	taacttcttc	ctgatggaat	ttattttctg	caagaattat	tctgatattt	101580

aagagagcca	attttaactg	ctgtgaaaat	gtttccagtg	caagagaagg	gaaatactag	101640
gaactaagac	atttctaatt	tattgcttat	tactttctta	attttacagg	ataattataa	101700
gcaagtggaa	ctaccatctt	ttattcttaa	taattattaa	tcccttcaat	gaaactttaa	101760
aaaaactgaa	tttttataca	tggcatacat	ttttctagtt	ccttctgctt	gctttattaa	101820
ctcaaaagtt	ctagttctag	tctggtgatc	tgccttttgt	tctcccaaaa	tgtacagtaa	101880
ttccatttgt	ttgtataaat	atgcctggat	tttcattata	aaaatgtcat	tgtaggaggat	101940
agagactcat	atcatggcct	tttaaataat	gtaataaagg	caaatagata	tttgccctta	102000
gtttactggg	taaaagtttg	tttacagaac	ttttctctgg	tgcttaaagt	atgctatgta	102060
aaatgtcatg	agtggaaaga	atatttgtag	tagtaacaag	aatttttcat	ttaggaaaga	102120
tttcttaggt	tttgaaagaa	tacattaaaa	taaaaaactt	gccctacta	ggtaagaact	102180
ttataatgaa	gacatacatt	cttcttaatt	ttactcttgc	tcttgttaaa	gatttgtttg	102240
aatatagaag	atgcatgatt	tctgggtttt	tttttttttt	tgagacagag	tttcgctctt	102300
gttgcccagg	ctggagtgca	atggcgcaat	ctcgactcac	cacaacctcc	gcctcccagg	102360
ttcaagcaat	tctcctgcct	cagcctcccg	agtagctggg	attacaggca	tgcgccacta	102420
ccccagctaa	ttttgtattt	ttagtagaga	tggggtttct	ccatgttggt	caggctggtc	102480
ttgaactcct	gacctcaggt	gatccgcctg	cctcggcctc	ccaaagtgct	gggattacag	102540
gcataagcca	ctgcgcccag	ccagaagatg	catgatttct	taggatcata	tgctgtttgt	102600
agccataagg	taaatcatgt	ctcttccaat	catgactttg	gaactccctg	aataataaaa	102660
atgagagttg	agataaatag	gggaaaaaaa	atttttttca	agccagagct	atgcatatgt	102720
taggtgatgg	gtagtatccc	tttaaggtct	caaacattac	aacatcaatt	atgaaatact	102780
gataacgaaa	ggtagtaatg	aaatatatat	gatgaaaaga	attgagaagt	tctaaattaa	102840
gacatttcag	ttaagctcat	aaaatttcat	tgttttcatt	taaaagatta	acgttattga	102900
tacttgagata	actgggcta	catattaaag	gactatgtgg	ttccagctca	acttttaata	102960
tattgtctcc	tttaaaacta	tcatggttat	aattctattg	ggaaagactt	ttagataaca	103020
aagatttcaa	atgttaaaaag	agataaaaagt	cagggttaata	ctatcttaaa	cactgagtca	103080
gaaaatcatt	actgtataga	agttgctttc	ctgatcaagt	ctgaacttca	gctagtgcta	103140
gagaactatt	ttctatgact	taactctaac	caagttttat	tttaagctgt	ttctttgata	103200
gaagggccat	gaaaatagag	taatgatata	gtaggagata	agggattggg	ttggtctttt	103260
tcaataaaga	tagaagttgc	tgaagttttc	tgaattaata	atgacttaga	ttgtgacctt	103320
ttagattcgg	tgttgagctc	tgtgttgat	tacttcctaa	aagataatgc	ttaaacatta	103380

agcattagt	tgctcttcat	gttaatatgg	cagagttttg	taaactaaat	taaaacttac	103440
tgatatattg	gactttgagc	caagggaaag	aatgagtact	atctttccag	atatcttaag	103500
ggtaaaagct	tattctaaga	cagtctgtcc	attgagaata	ttagatttct	gacttgcaaa	103560
tatgtttgta	ctccagaaga	attagaggaa	aagcagatac	tagaattcta	atttaattac	103620
atatacagcc	gtctttgttt	atagtgtaga	attctttata	ttttgtacaa	aaactaattc	103680
ttttggtaaa	atgaaccatt	tacagttcgg	ttttggactc	tgagtcaaag	gattttcctt	103740
taaatgcttg	tctcaatttt	agtctggtct	tttgactttt	tcttcagaag	aaatgaatta	103800
aagggtacag	ttgcataaag	tgggttttta	tcctaatagt	ttggaaataa	atgataaact	103860
ttattttgtc	tttacttttt	tacttttaaaa	ctttttgata	ttttaggggt	tggagtctga	103920
taatgaagga	gttgtgtgta	ttggactctt	agtaacaatt	ataaacgctt	aacaaaatat	103980
agaaagcaat	gattgggtgg	ctctgcagag	caatcaaaac	aaggtagaaa	ctgcaaagtc	104040
ctataatgga	agagattcag	gctgggtgtg	gtggctcatg	cctgtttggg	aggccagtat	104100
gggagaatcg	cttacctca	ggaattcaag	accagcctgg	ggaatatagg	gagaccctgt	104160
ttctgccaaa	aaaaaaaaaa	aaaaaaaaaac	taaactaaaa	attagccagg	tttggtggcc	104220
tacacctgta	gtcccggtta	ctcaagaggc	tgaggttaga	ggatcgctta	agcccgggag	104280
gaggaggttg	cagtgaacca	agatcacacc	accacactcc	agcccagggt	acagagaacc	104340
tgtctcaaaa	aataaaaagt	aaacgagatt	caccttaacc	agcttttacc	catagggcaa	104400
tttcagttt	gtgcagttca	cttgatata	gaattcaggt	agaatgtgac	agttttgaca	104460
ggctgaggaa	tcatttgga	cttctagaac	acctgaaaat	tagaagagaa	attttgaggaa	104520
ggagagtgcc	tcagaaagta	agctctaaaa	tctgcctaca	tattcctttc	aatccttgg	104580
tggattccaa	aattgtgctg	gtgcagtgtg	attatttaaa	ggaaccagag	agaaagcaag	104640
ttaaagtcta	aaaaactcag	cagcgattgc	agctgcccaa	ggacagagag	tttgaaattc	104700
aagttccact	aagaaggact	gagtaaatac	ttggtgtttc	ccattgaacc	ccaaaatgc	104760
cacgccttaa	ggaagaatga	cagcatccta	gaactaaagg	ctgtgcttca	gcactaagga	104820
caaatggaa	ataaacttac	tgtaatgaag	cttaaaatca	aatctcacag	catcaagggt	104880
atccattagt	aattttaaac	cagaacaaaa	cacagcattc	ttgagaagaa	aacaattcag	104940
tcatcacagc	atatatccaa	agctcagtac	ataataaaca	gttactaagc	atgcaaagaa	105000
aaattatgtg	atccttttga	ctcaacttcc	ctgatggaaa	aaaaaaagtg	gtccatacca	105060
tactcaacag	aaaatatata	gaagcagatg	cacaagtgac	ctctatgaga	gtgattttaa	105120
aatagcggtta	aaatatgtta	aaagaattta	caagaagaga	taatgggtga	atagataggg	105180

agtttcagca	gaaaaataag	aattgaaaac	ccaacagagg	ccaggcaccg	tgcttcacgc	105240
ccgtaatcac	agcactttgg	gaggccaagg	caggtggatc	acctaaggtc	aggagttcaa	105300
gaccagcctg	accaacatgg	tgaaacccgg	tctctactaa	aaatacaaaa	ctagccgggc	105360
atgggtgatg	catgcccata	atcctagcta	cttgggaggg	tgaagcagga	gaattgcttg	105420
aacctgggag	gtagagggtg	cagtgaacca	agatcatgcc	attgcactcc	agcctgggca	105480
agaagagcaa	aactccatct	caaaaaataa	taataataat	tggctgggcg	cgggtggctca	105540
cgcctgtaat	cccagcactt	tgggaggccg	agacaggcag	atcacaaggt	caggagattg	105600
agaccatcct	ggctaacatg	gtgaaacccc	gtctctacta	aaaatacaaa	aaattagcct	105660
ggcatggtgg	cgggcacctg	tagtcccagc	tactcgggag	gctgaagcag	gagaatggca	105720
tgaccccagg	aggcggaact	tgcaatgagc	ccagatcgcg	ccactgcact	ccagcctggg	105780
caaccgtctc	aaaaataata	ataataataa	ttagaaaata	ataaaaaata	accactggga	105840
aagtcttgta	atgaaaacta	cagtatctga	aattaaaaat	ttaaatggat	ggacttaaca	105900
gtaaactagg	cacaacagga	aagactgaaa	aagacaagta	agtatcaaac	caaagcagag	105960
agagaaagaa	aaaagaaata	gcagaggttt	gagacctttg	gaacattatc	agttcctgta	106020
at ttgaagta	aagaaggaaa	ggaagagaat	ggggcaagag	aaatctttga	agcaatgatg	106080
cgtgaaat tt	tcccaaagt	gttgaaagac	atcaacgtac	agatctaaga	agttcactga	106140
atccccaagc	agaataaata	caacaaagac	tacatctggg	tacatcacat	tgccaaaaaa	106200
aaaaaaaaaa	aaaatcttaa	aagcaatcaa	ggtggggagt	gggaaggagt	agcattacat	106260
tcagataaac	aagagtgatg	gctaacctca	tcagaaatga	tggaaattaag	aaaatgatgg	106320
aatgacatct	ttaagggtgca	aagatgaaac	aaagggttaac	ttaaaactct	gtatccagtt	106380
aaaatatcct	tcaagatggt	aaggcaatgg	ccaggctcag	tggcttatgc	ctgtaatccc	106440
aacactttgg	ggggctgaga	caggagaatc	ccttgaggcc	atgagtgacc	agcctgggca	106500
acctagtgg	aaccccatca	aaaaataaaa	ttagctgggc	atgggctgtg	gtcgtgctac	106560
tgactccag	cctgggggac	agagtcctta	tctcttgaaa	aacttaagac	aaaatctttt	106620
ccagatgaaa	acagaatttg	ttgctggcag	actgacttta	tacggaatgc	tttaagaaat	106680
tcttcaggct	gaaggaaaat	ataaatctgg	agtgaaaaac	gtagtatctg	aaattaaaaat	106740
tttccttgga	tggacttatc	agtaaatgga	tttcccaatg	ggaacctgta	tctccaggat	106800
ggaattaaga	at tttgaaat	gataaacatg	ttggtaaata	agaaagacat	ttttcttttt	106860
tattttttcta	aatatcattg	ctgttttaag	caaacaaaaa	tagtattgca	gtgtttataa	106920
tatatgtata	agtaaaatat	aggacagaaa	tagcatgaag	ggtcagagag	ggcaataaat	106980

gtaattatat	tgagataata	ttactatatt	gtacacaaag	tgagatagca	ttagtgtaag	107040
gtagattgtg	ataagaatgc	atdddgtaat	ctctcacccc	atcactgggg	tgcagtagat	107100
ctaaaaagcc	attagagaag	ataaaatggg	tcactaagaa	ttatgtgatt	agtcaaataga	107160
aatcaaggaa	agaaaaacag	agcaaagaag	tgtaagataa	aacagtggga	tggaagagct	107220
gaaattaata	atgtcactaa	attgcttacc	ttcctccttc	ctgaactgtt	gtggattttt	107280
tttttttcag	gggggtgagg	ttttttttta	gagacagagt	ctctattgcc	ccaggctgga	107340
cagcagtggc	gtaatcagtc	tcaaaatccc	tgcaagtctca	aactcctggg	ctcaagtgat	107400
cctcctgcct	cggcctccca	aaatgctggg	attacaggca	ggcagagcca	ccacaccctg	107460
ccagaactat	tgtttttaaag	tcattaggtg	gttaagccct	cgagactatg	tgcaagtttg	107520
atgattcggt	aggagcactc	acaggactca	gcatatagtc	ttattcccaa	caatgattta	107580
ttagagcaaa	aggatatgaa	gcaaattcag	caaagggaag	aaaggaatta	agtgaagtcc	107640
agaggaaacc	aagcaccagc	ttctaagggt	tctctcctaa	tgaagtcaca	caggatgtca	107700
ctccagcaac	aagtgcacgc	atgtgaaagt	gttaagcctc	actgttagac	atagtaagct	107760
tgtagaggaa	tagagggaat	ctcccctcaa	attcaagttc	ccagatgcct	ggcaaggggc	107820
aaccttgcca	gcaggacttt	ctggggatag	cagtcccagg	cacgtgtttg	cacagtgggt	107880
cagatcaatg	tgcttacatt	gggtagaggg	gacctatgga	agtccaaatt	tgggtgtcag	107940
agacctataa	gggtgaagac	agtgtctaca	atgatggaca	gctaggtatg	aggtgtcagg	108000
gacagattga	ggcagttatt	cacatgggtg	ataggggcaa	cctggaatga	ggaataagct	108060
caagcacaga	aaggggtgtc	catatggggg	aaggggtgtc	gcagagatgg	tagattgggt	108120
gcacacaaaa	gacttgatgg	aataaagtga	atataactaac	cacagagaag	gtaattataa	108180
atacagaaaag	gggggaaact	agaatgaaac	ctggagttca	gcttgaattg	ggtttaagaa	108240
agtgaattca	tggttttaaaa	tctataaaaa	tagatgaaat	atcggtggg	cacagtggct	108300
cacgcatata	atcccagcac	tttgggaggc	cgaggcgggc	agatcgcttg	agatcaggag	108360
tttgagacca	gcctggccaa	cactgcgaaa	ccccatctct	agtaaaaata	caaaaattag	108420
ccagggtgtg	tggtacacac	ctgtagtccc	agctacttgg	gaggctgagg	cacgagaatc	108480
acttgaacct	gggaggcaga	ggttgcaagt	ggctgagatc	gcaccactgc	actccagcct	108540
gggcaacaga	gcgagatcct	gcctcaaata	aaaatagagg	acatataagc	atataaatat	108600
acacgtgtgt	atgtgtccaa	atatgtatat	tccctagtct	gtccaccaag	gtggccttgg	108660
agcagttatg	ctccaataat	aatgagcaca	taaagtaccc	atatcttgcc	ttccaaattc	108720
ttcactgtct	tagtctgctt	gggctgcatt	acaaaatacc	atagactggg	cagcttaaat	108780

aacagaaatt	tatcctcaca	gttctagaag	cttggaagtc	caagattaaa	gtaccagcca	108840
gttttggttc	tagtgagggc	tttcttcctg	gcttgagat	ggccaccttc	tcaccgtgtc	108900
cttgtagggc	agacagcaca	agctctctgg	tgtccctttt	taaaagggca	ttaatcccat	108960
catgacagtc	ccatcctcat	tatctcatct	aaccctaggt	acttcccaaa	ggctgaatca	109020
ccaaagacca	tcacattgct	ggtgaaggct	tcaacatatg	aatttgaggg	acacgaatat	109080
tcagtccata	acatcaacta	aaggaaccaa	gactctttga	taaaatggct	aaattcaggg	109140
ctggggcaga	gaaaatacat	gagtgtggaa	cttcttgtgc	cagagagaaa	aagtgcccaa	109200
agattgatga	ggatgaatca	ttgaaatgac	acacagatta	aaagggttcc	cactggacaa	109260
atttgagcat	caaaataagt	aatagtagta	attaattata	acccatcaga	agaaataaac	109320
catgagctca	tgtgaatata	tgaatacaaa	cataaaca	ttacaagcat	aatgaggaat	109380
gtgatattta	tatggtttta	aggtacctct	ccaggccggg	tgcagtaact	ctcacctgta	109440
atcccagcac	tttgggaggg	caaggcaggt	agatcacctg	aggtcaggca	tttgagacca	109500
gcctgcacaa	catggtgaaa	ccctgactct	actaaaaata	cataacgcga	gccgggctgt	109560
gtggcacgtg	tctataatct	gccactgatt	aggtgtgtga	ttttcccaag	caggggataa	109620
tagtagtacc	tatgtcaaag	gctgttatga	ggattaaatg	agctaacaca	taatcgtgct	109680
tttttttttt	tttttttttt	ttgagacaga	gtcttgact	gtcgctggg	ctggagtgca	109740
atggcacgat	ctcgccccc	tgcaacctct	gcctcccagg	ttcaagtgat	tctcctgcct	109800
cagcctcctg	agtagctggg	attacaggct	cctgccacca	cacctggcta	ttttcaatag	109860
agacgggggt	tcactatggt	ggccaggcta	gtctcaaaaa	cctgacctcg	tgatccaccc	109920
gctttggcct	cccaaagtgc	tgggattaca	ggcatgagcc	actgcacccg	gctttttttt	109980
tttttttttg	agatggaatc					110000

<210> 2

<211> 3263

<212> DNA

<213> Homo sapiens

<400> 2

gtcctgaga	ccggcgggca	cacgggggtc	tgtggccccc	gccgtagcag	tggctgccgc	60
cgtcgcttgg	ttcccgtcgg	tctgctggag	gcgggttatg	gcggcgccgg	cagtgagagc	120
tgtgaatgaa	ttctccgggt	ggacgaggga	agaagaaagg	ctccggcggc	gccagcaacc	180
cggtgcctcc	caggcctccg	cccccttgcc	tggccccgc	ccctcccgcc	gccgggcccg	240
cccctccgcc	cgagtcgccg	cataagcgga	acctgtacta	tttctcctac	ccgctgtttg	300
taggcttcgc	gctgctgcgt	ttggctgcct	tccacctggg	gctcctcttc	gtgtggctct	360

gccagcgctt	ctcccgcgcc	ctcatggcag	ccaagaggag	ctccggggcc	gcgccagcac	420
ctgcctcggc	ctcgccccg	gcgcgggtgc	cgggcggcga	ggccgagcgc	gtccgagtct	480
tccacaaaca	ggccttcgag	tacatctcca	ttgccctgcg	catcgatgag	gatgagaaag	540
caggacagaa	ggagcaagct	gtggaatggt	ataagaaagg	tattgaagaa	ctggaaaaag	600
gaatagctgt	tatagttaca	ggacaagggtg	aacagtgtga	aagagctaga	cgccttcaag	660
ctaaaatgat	gactaatttg	gttatggcca	aggaccgctt	acaacttcta	gagaagatgc	720
aaccagtttt	gccattttcc	aagtcacaaa	cggacgtcta	taatgacagt	actaacttgg	780
catgccgcaa	tggacatctc	cagtcagaaa	gtggagctgt	tccaaaaaga	aaagaccctt	840
taacacacac	tagtaattca	ctgcctcggt	caaaaacagt	tatgaaaact	ggatctgcag	900
gcctttcagg	ccaccataga	gcacctagtt	acagtggttt	atccatgggt	tctggagtga	960
aacagggatc	tggctctgct	cctaccactc	ataagggtag	tccgaaaaca	aataggacaa	1020
ataaaccttc	tacccttaca	actgctactc	gtaagaaaaa	agacttgaag	aatttttagga	1080
atgtggacag	caaccttgct	aaccttataa	tgaatgaaat	tgtggacaat	ggaacagctg	1140
ttaaatttga	tgatatagct	ggtcaagact	tggcaaaaaca	agcattgcaa	gaaattgtta	1200
ttcttccttc	tctgaggcct	gagttgttca	cagggcttag	agctcctgcc	agagggctgt	1260
tactctttgg	tccacctggg	aatgggaaga	caatgctggc	taaagcagta	gctgcagaat	1320
cgaatgcaac	cttctttaat	ataagtgtcg	caagtttaac	ttcaaaatac	gtggggagaag	1380
gagagaaatt	ggtgagggct	ctttttgctg	tggctcgaga	acttcaacct	tctataattt	1440
ttatagatga	agttgatagc	cttttgctgtg	aaagaagaga	aggggagcac	gatgctagta	1500
gacgcctaaa	aactgaattt	ctaatagaat	ttgatgggtg	acagtctgct	ggagatgaca	1560
gagtacttgt	aatgggtgca	actaataggc	cacaagagct	tgatgaggct	gttctcaggc	1620
gtttcatcaa	acgggtatat	gtgtctttac	caaatgagga	gacaagacta	cttttgctta	1680
aaaatctggt	atgtaaacaa	ggaagtccat	tgacccaaaa	agaactagca	caacttgcta	1740
gaatgactga	tggatactca	ggaagtgacc	taacagcttt	ggcaaaagat	gcagcactgg	1800
gtcctatccg	agaactaaaa	ccagaacagg	tgaagaatat	gtctgccagt	gagatgagaa	1860
atattcgatt	atctgacttc	actgaatcct	tgaaaaaaat	aaaacgcagc	gtcagccctc	1920
aaactttaga	agcgtacata	cgttggaaca	aggactttgg	agataccact	gtttaaggaa	1980
atacctttgt	aaacctgcag	aacattttac	ttaaaagagg	aaacacaaga	tcttcaatga	2040
acgtcatcgg	ctacagaaac	agcctaagtt	tacaggactt	tttagagtct	tacatatattg	2100
tgcaccaaac	ttgaagatga	accagaaaac	agacttaaac	aaaatataca	atgcaaatgt	2160

```

aattttttgt  tgtttaaggc  cttgccttga  tggtcacagt  tatcccaatg  gacactaagt  2220
tagagcacaa  caaaacctga  ttctgggtctt  ctttaccaat  ataatcataa  tgtaaataat  2280
aatttgtata  ttgtgttgca  gatgaaagta  ttccaggaa  agtgaatggt  agaagacaca  2340
agaacatttg  tttgtttgtc  ttctgatgtt  ttttcttaaa  atagtaattt  ctctactttt  2400
tcttttctac  tgttgtctta  actacagggtg  attggaatgc  caaacactct  taagtttatt  2460
ttcttttttc  gttttataaa  ttcagtgtgc  caaatgaaac  ttttttccta  agtaactgta  2520
ataggaaaaa  gtttatattg  agagtttctt  cttcataaat  ctacagacat  taaacaattg  2580
ttgtgttctt  tttacctttt  atttttctat  taccttgcta  ccaaacagtt  tagatagcaa  2640
tataatagca  aaaaagcaaa  tatggtaaaa  tagagaaggt  ttgaaggttt  gagttactct  2700
gtcatataac  atgtagatca  gtcttcatgt  gacctgcagt  attttttttt  ctaatgtatt  2760
tgtcagaaat  ctgtttaga  ctgttaactt  cttcctgatg  gaatttattt  tctgcaagaa  2820
ttattctgat  atttaagaga  gccaatttta  actgctgtga  aaatgtttcc  agtgcaagag  2880
aagggaata  ctaggaacta  agacatttct  aatttattgc  ttattacttt  cttaatttta  2940
caggataatt  ataagcaagt  ggaactacca  tcttttattc  ttaataatta  ttaatccctt  3000
caatgaaact  ttaaaaaaac  tgaattttta  tacatggcat  acatttttct  agttccttct  3060
gcttgcttta  ttaactcaaa  agttctagtt  ctagtctgtt  gatctgcctt  ttgttctccc  3120
aaaatgtaca  gtaattccat  ttgtttgtat  aaatatgcct  ggattttcat  tataaaaatg  3180
tcattgtagg  gagtagagac  tcatatcatg  gccttttaaa  tattgtaata  aaggcaaata  3240
gatatttgcc  cttagtttac  tgg  3263

```

<210> 3

<211> 616

<212> PRT

<213> Homo sapiens

<400> 3

```

Met Asn Ser Pro Gly Gly Arg Gly Lys Lys Lys Gly Ser Gly Gly Ala
 1              5              10              15

Ser Asn Pro Val Pro Pro Arg Pro Pro Pro Pro Cys Leu Ala Pro Ala
              20              25              30

Pro Pro Ala Ala Gly Pro Ala Pro Pro Pro Glu Ser Pro His Lys Arg
              35              40              45

Asn Leu Tyr Tyr Phe Ser Tyr Pro Leu Phe Val Gly Phe Ala Leu Leu
 50              55              60

Arg Leu Val Ala Phe His Leu Gly Leu Leu Phe Val Trp Leu Cys Gln
 65              70              75              80

```

Arg	Phe	Ser	Arg	Ala	Leu	Met	Ala	Ala	Lys	Arg	Ser	Ser	Gly	Ala	Ala	85	90	95
Pro	Ala	Pro	Ala	Ser	Ala	Ser	Ala	Pro	Ala	Pro	Val	Pro	Gly	Gly	Glu	100	105	110
Ala	Glu	Arg	Val	Arg	Val	Phe	His	Lys	Gln	Ala	Phe	Glu	Tyr	Ile	Ser	115	120	125
Ile	Ala	Leu	Arg	Ile	Asp	Glu	Asp	Glu	Lys	Ala	Gly	Gln	Lys	Glu	Gln	130	135	140
Ala	Val	Glu	Trp	Tyr	Lys	Lys	Gly	Ile	Glu	Glu	Leu	Glu	Lys	Gly	Ile	145	150	155
Ala	Val	Ile	Val	Thr	Gly	Gln	Gly	Glu	Gln	Cys	Glu	Arg	Ala	Arg	Arg	165	170	175
Leu	Gln	Ala	Lys	Met	Met	Thr	Asn	Leu	Val	Met	Ala	Lys	Asp	Arg	Leu	180	185	190
Gln	Leu	Leu	Glu	Lys	Met	Gln	Pro	Val	Leu	Pro	Phe	Ser	Lys	Ser	Gln	195	200	205
Thr	Asp	Val	Tyr	Asn	Asp	Ser	Thr	Asn	Leu	Ala	Cys	Arg	Asn	Gly	His	210	215	220
Leu	Gln	Ser	Glu	Ser	Gly	Ala	Val	Pro	Lys	Arg	Lys	Asp	Pro	Leu	Thr	225	230	235
His	Thr	Ser	Asn	Ser	Leu	Pro	Arg	Ser	Lys	Thr	Val	Met	Lys	Thr	Gly	245	250	255
Ser	Ala	Gly	Leu	Ser	Gly	His	His	Arg	Ala	Pro	Ser	Tyr	Ser	Gly	Leu	260	265	270
Ser	Met	Val	Ser	Gly	Val	Lys	Gln	Gly	Ser	Gly	Pro	Ala	Pro	Thr	Thr	275	280	285
His	Lys	Gly	Thr	Pro	Lys	Thr	Asn	Arg	Thr	Asn	Lys	Pro	Ser	Thr	Pro	290	295	300
Thr	Thr	Ala	Thr	Arg	Lys	Lys	Lys	Asp	Leu	Lys	Asn	Phe	Arg	Asn	Val	305	310	315
Asp	Ser	Asn	Leu	Ala	Asn	Leu	Ile	Met	Asn	Glu	Ile	Val	Asp	Asn	Gly	325	330	335
Thr	Ala	Val	Lys	Phe	Asp	Asp	Ile	Ala	Gly	Gln	Asp	Leu	Ala	Lys	Gln	340	345	350
Ala	Leu	Gln	Glu	Ile	Val	Ile	Leu	Pro	Ser	Leu	Arg	Pro	Glu	Leu	Phe	355	360	365
Thr	Gly	Leu	Arg	Ala	Pro	Ala	Arg	Gly	Leu	Leu	Leu	Phe	Gly	Pro	Pro	370	375	380
Gly	Asn	Gly	Lys	Thr	Met	Leu	Ala	Lys	Ala	Val	Ala	Ala	Glu	Ser	Asn	385	390	395
																		400

Ala Thr Phe Phe Asn Ile Ser Ala Ala Ser Leu Thr Ser Lys Tyr Val
405 410 415

Gly Glu Gly Glu Lys Leu Val Arg Ala Leu Phe Ala Val Ala Arg Glu
420 425 430

Leu Gln Pro Ser Ile Ile Phe Ile Asp Glu Val Asp Ser Leu Leu Cys
435 440 445

Glu Arg Arg Glu Gly Glu His Asp Ala Ser Arg Arg Leu Lys Thr Glu
450 455 460

Phe Leu Ile Glu Phe Asp Gly Val Gln Ser Ala Gly Asp Asp Arg Val
465 470 475 480

Leu Val Met Gly Ala Thr Asn Arg Pro Gln Glu Leu Asp Glu Ala Val
485 490 495

Leu Arg Arg Phe Ile Lys Arg Val Tyr Val Ser Leu Pro Asn Glu Glu
500 505 510

Thr Arg Leu Leu Leu Leu Lys Asn Leu Leu Cys Lys Gln Gly Ser Pro
515 520 525

Leu Thr Gln Lys Glu Leu Ala Gln Leu Ala Arg Met Thr Asp Gly Tyr
530 535 540

Ser Gly Ser Asp Leu Thr Ala Leu Ala Lys Asp Ala Ala Leu Gly Pro
545 550 555 560

Ile Arg Glu Leu Lys Pro Glu Gln Val Lys Asn Met Ser Ala Ser Glu
565 570 575

Met Arg Asn Ile Arg Leu Ser Asp Phe Thr Glu Ser Leu Lys Lys Ile
580 585 590

Lys Arg Ser Val Ser Pro Gln Thr Leu Glu Ala Tyr Ile Arg Trp Asn
595 600 605

Lys Asp Phe Gly Asp Thr Thr Val
610 615

<210> 4
<211> 23
<212> ADN
<213> Artificial Sequence

<220>
<223> Amorce

<400> 4
cggagctcct cttggctgcc atg

23

<210> 5
<211> 26
<212> ADN
<213> Artificial Sequence

<220>

<223> Amorce

<400> 5

agaagcgctg gcagagccac acgaag

26

<210> 6

<211> 27

<212> ADN

<213> Artificial Sequence

<220>

<223> Amorce

<400> 6

aaggcgacca aacgcagcag cgcgaag

27

<210> 7

<211> 26

<212> ADN

<213> Artificial Sequence

<220>

<223> Amorce

<400> 7

aggagcaagc tgtggaatgg tataag

26

<210> 8

<211> 27

<212> ADN

<213> Artificial Sequence

<220>

<223> Amorce

<400> 8

tggttatggc caaggaccgc ttacaac

27

<210> 9

<211> 26

<212> ADN

<213> Artificial Sequence

<220>

<223> Amorce

<400> 9

caaacggacg tctataatga cagtac

26

<210> 10

<211> 25

<212> ADN

<213> Artificial Sequence

<220>

<223> Amorce

<400> 10

ttaggaatgt ggacagcaac cttgc

25

<210> 11
 <211> 25
 <212> ADN
 <213> Artificial Sequence

<220>
 <223> Amorce

<400> 11
 cttctctgag gcctgagttg ttcac

25

<210> 12
 <211> 27
 <212> ADN
 <213> Artificial Sequence

<220>
 <223> Amorce

<400> 12
 tgctagaatg actgatggat actcagg

27

<210> 13
 <211> 24
 <212> ADN
 <213> Artificial Sequence

<220>
 <223> Amorce

<400> 13
 agatgcagca ctgggtccta tccg

24

<210> 14
 <211> 26
 <212> ADN
 <213> Artificial Sequence

<220>
 <223> Amorce

<400> 14
 atgaacgtca tcggctacag aaacag

26

<210> 15
 <211> 19
 <212> ADN
 <213> Artificial Sequence

<220>
 <223> Amorce

<400> 15
 tagcagtggc tgccgccgt

19

<210> 16
 <211> 20
 <212> ADN
 <213> Artificial Sequence

<220>
 <223> Amorce

 <400> 16
 aagcggtcct tggccataac 20

 <210> 17
 <211> 19
 <212> ADN
 <213> Artificial Sequence

 <220>
 <223> Amorce

 <400> 17
 ggcggcagtg agagctgtg 19

 <210> 18
 <211> 20
 <212> ADN
 <213> Artificial Sequence

 <220>
 <223> Amorce

 <400> 18
 ctagctcttt cacactgttc 20

 <210> 19
 <211> 20
 <212> ADN
 <213> Artificial Sequence

 <220>
 <223> Amorce

 <400> 19
 aacaggcctt cgagtacatc 20

 <210> 20
 <211> 20
 <212> ADN
 <213> Artificial Sequence

 <220>
 <223> Amorce

 <400> 20
 ctgtgaacaa ctcaggcctc 20

 <210> 21
 <211> 20
 <212> ADN
 <213> Artificial Sequence

 <220>
 <223> Amorce

 <400> 21

atgagaaagc aggacagaag	20
<210> 22	
<211> 18	
<212> ADN	
<213> Artificial Sequence	
<220>	
<223> Amorce	
<400> 22	
tgccaagtct tgaccagc	18
<210> 23	
<211> 20	
<212> ADN	
<213> Artificial Sequence	
<220>	
<223> Amorce	
<400> 23	
ctacaactgc tactcgtaag	20
<210> 24	
<211> 19	
<212> ADN	
<213> Artificial Sequence	
<220>	
<223> Amorce	
<400> 24	
cagtgctgca tcttttgcc	19
<210> 25	
<211> 20	
<212> ADN	
<213> Artificial Sequence	
<220>	
<223> Amorce	
<400> 25	
taggaatgtg gacagcaacc	20
<210> 26	
<211> 20	
<212> ADN	
<213> Artificial Sequence	
<220>	
<223> Amorce	
<400> 26	
aaagctgtta ggtcacttcc	20
<210> 27	
<211> 20	

<212> ADN
 <213> Artificial Sequence

<220>
 <223> Amorce

<400> 27
 tggagatgac agagtacttg 20

<210> 28
 <211> 19
 <212> ADN
 <213> Artificial Sequence

<220>
 <223> Amorce

<400> 28
 ctggaataact ttcactctgc 19

<210> 29
 <211> 19
 <212> ADN
 <213> Artificial Sequence

<220>
 <223> Amorce

<400> 29
 atgaggctgt tctcaggcg 19

<210> 30
 <211> 20
 <212> ADN
 <213> Artificial Sequence

<220>
 <223> Amorce

<400> 30
 gtgagccgaa ctgcacattg 20

<210> 31
 <211> 21
 <212> ADN
 <213> Artificial Sequence

<220>
 <223> Amorce

<400> 31
 caaagtcgac agctacagtg c 21

<210> 32
 <211> 20
 <212> ADN
 <213> Artificial Sequence

<220>

<223> Amorce

<400> 32

ggaactgtag ttgagtggga

20

<210> 33

<211> 19

<212> ADN

<213> Artificial Sequence

<220>

<223> Amorce

<400> 33

agatgaggct ccgacctac

19

<210> 34

<211> 20

<212> ADN

<213> Artificial Sequence

<220>

<223> Amorce

<400> 34

aatgccacac ttgtaatctc

20

<210> 35

<211> 22

<212> ADN

<213> Artificial Sequence

<220>

<223> Amorce

<400> 35

tgtgaatata tcataatttg gg

22

<210> 36

<211> 19

<212> ADN

<213> Artificial Sequence

<220>

<223> Amorce

<400> 36

tacagcagtt ctcacgatg

19

<210> 37

<211> 21

<212> ADN

<213> Artificial Sequence

<220>

<223> Amorce

<400> 37

gaccaaattg gtgcatgcat g

21

<210> 38
 <211> 20
 <212> ADN
 <213> Artificial Sequence

<220>
 <223> Amorce

<400> 38
 acattttccaa tacatcccac 20

<210> 39
 <211> 20
 <212> ADN
 <213> Artificial Sequence

<220>
 <223> Amorce

<400> 39
 atttgtcatt tcacatgcac 20

<210> 40
 <211> 20
 <212> ADN
 <213> Artificial Sequence

<220>
 <223> Amorce

<400> 40
 ttagaatgac tatacctgac 20

<210> 41
 <211> 18
 <212> ADN
 <213> Artificial Sequence

<220>
 <223> Amorce

<400> 41
 tcagggttaag taagactc 18

<210> 42
 <211> 19
 <212> ADN
 <213> Artificial Sequence

<220>
 <223> Amorce

<400> 42
 ttcctatcta cctagtgac 19

<210> 43
 <211> 20
 <212> ADN
 <213> Artificial Sequence

<220>
 <223> Amorce

<400> 43
 ttttatagca agttgccctg 20

<210> 44
 <211> 19
 <212> ADN
 <213> Artificial Sequence

<220>
 <223> Amorce

<400> 44
 cctatgaaga tcctggtac 19

<210> 45
 <211> 20
 <212> ADN
 <213> Artificial Sequence

<220>
 <223> Amorce

<400> 45
 tgtcatgatt ctaacaagg 20

<210> 46
 <211> 20
 <212> ADN
 <213> Artificial Sequence

<220>
 <223> Amorce

<400> 46
 tctatttcac tcctgacatg 20

<210> 47
 <211> 19
 <212> ADN
 <213> Artificial Sequence

<220>
 <223> Amorce

<400> 47
 gtcataagggc ttaggcttc 19

<210> 48
 <211> 20
 <212> ADN
 <213> Artificial Sequence

<220>
 <223> Amorce

<400> 48

atcatactac ccacttttcc	20
<210> 49	
<211> 20	
<212> ADN	
<213> Artificial Sequence	
<220>	
<223> Amorce	
<400> 49	
tgtttgaggaa gatgctactg	20
<210> 50	
<211> 21	
<212> ADN	
<213> Artificial Sequence	
<220>	
<223> Amorce	
<400> 50	
ctactgaaga taacgtacat g	21
<210> 51	
<211> 20	
<212> ADN	
<213> Artificial Sequence	
<220>	
<223> Amorce	
<400> 51	
cattgattgc catgtattgg	20
<210> 52	
<211> 20	
<212> ADN	
<213> Artificial Sequence	
<220>	
<223> Amorce	
<400> 52	
agaaggccag aaatactcag	20
<210> 53	
<211> 22	
<212> ADN	
<213> Artificial Sequence	
<220>	
<223> Amorce	
<400> 53	
gtacttaaataat cggtaaatat gg	22
<210> 54	
<211> 20	

<212> ADN
 <213> Artificial Sequence

<220>
 <223> Amorce

<400> 54
 ctcaagtctt aggaatgcag 20

<210> 55
 <211> 20
 <212> ADN
 <213> Artificial Sequence

<220>
 <223> Amorce

<400> 55
 gcacttaacc aggctgtatg 20

<210> 56
 <211> 19
 <212> ADN
 <213> Artificial Sequence

<220>
 <223> Amorce

<400> 56
 ctcagatgac tcacatagc 19

<210> 57
 <211> 22
 <212> ADN
 <213> Artificial Sequence

<220>
 <223> Amorce

<400> 57
 ctttactaga ctaattctcc tg 22

<210> 58
 <211> 20
 <212> ADN
 <213> Artificial Sequence

<220>
 <223> Amorce

<400> 58
 cagattcaag aagacagatc 20

<210> 59
 <211> 20
 <212> ADN
 <213> Artificial Sequence

<220>

<223> Amorce

<400> 59

gcaataattc accacacttg

20

<210> 60

<211> 20

<212> ADN

<213> Artificial Sequence

<220>

<223> Amorce

<400> 60

ggtagttctt gtttctgctc

20

<210> 61

<211> 20

<212> ADN

<213> Artificial Sequence

<220>

<223> Amorce

<400> 61

caagtgtggt gaattattgc

20

<210> 62

<211> 19

<212> ADN

<213> Artificial Sequence

<220>

<223> Amorce

<400> 62

gagctgaaaa gtattcagc

19

<210> 63

<211> 20

<212> ADN

<213> Artificial Sequence

<220>

<223> Amorce

<400> 63

tgcaaaggac atagccagtg

20

<210> 64

<211> 20

<212> ADN

<213> Artificial Sequence

<220>

<223> Amorce

<400> 64

agcctctgga gatagtatgc

20

<210> 65
 <211> 20
 <212> ADN
 <213> Artificial Sequence

<220>
 <223> Amorce

<400> 65
 ctagaacagg ggtcacagtc 20

<210> 66
 <211> 18
 <212> ADN
 <213> Artificial Sequence

<220>
 <223> Amorce

<400> 66
 ttggacttct taaacttc 18

<210> 67
 <211> 21
 <212> ADN
 <213> Artificial Sequence

<220>
 <223> Amorce

<400> 67
 gcagtatgca agaaattgaa c 21

<210> 68
 <211> 19
 <212> ADN
 <213> Artificial Sequence

<220>
 <223> Amorce

<400> 68
 ggcctgtaat tttcttctg 19

<210> 69
 <211> 21
 <212> ADN
 <213> Artificial Sequence

<220>
 <223> Amorce

<400> 69
 gtactgaata gatacatgta g 21

<210> 70
 <211> 19
 <212> ADN
 <213> Artificial Sequence

<220>

<223> Amorce

<400> 70

gtgtagcaga tcaacatag

19

<210> 71

<211> 20

<212> ADN

<213> Artificial Sequence

<220>

<223> Amorce

<400> 71

catcttcaag tttggtgcac

20

REVENDECATIONS

1. Acide nucléique purifié ou isolé du gène SPG4, caractérisé en ce qu'il comprend au moins 15 nucléotides consécutifs d'une séquence choisie parmi le
5 groupe comprenant la séquence SEQ ID No. 1, les séquences nucléiques homologues ou variantes de l'acide nucléique de séquence SEQ ID No. 1, leur séquence complémentaire et la séquence de leur ARN correspondant.
2. Acide nucléique purifié ou isolé selon la revendication 1, caractérisé en ce qu'il comprend une séquence choisie parmi le groupe comprenant la séquence SEQ
10 ID No. 1, la séquence SEQ ID No. 2, la séquence SEQ ID No. 72, les séquences nucléiques homologues ou variantes des séquences SEQ ID No. 1, SEQ ID No. 2 ou SEQ ID No. 72, leur séquence complémentaire et la séquence de leur ARN correspondant.
3. Acide nucléique purifié ou isolé selon la revendication 1 ou 2,
15 caractérisé en ce qu'il comprend au moins une mutation dont la position et la nature sont identifiées au tableau 5.
4. Sonde ou amorce, caractérisée en ce qu'elle comprend une séquence d'un acide nucléique selon l'une des revendications 1 à 3.
5. Sonde ou amorce selon la revendication 4, caractérisée en ce que sa
20 séquence est choisie parmi les séquences SEQ ID No. 4 à SEQ ID No. 71.
6. Site accepteur ou donneur d'épissage, caractérisé en ce que sa séquence est choisie parmi les séquences de sites accepteurs ou donneurs d'épissage identifiées au tableau 3.
7. Méthode pour le criblage de banques d'ADNc ou d'ADN génomique ou
25 pour le clonage d'ADNc ou génomique isolé codant pour la spastin, caractérisée en ce qu'elle met en œuvre une séquence nucléique selon l'une des revendications 1 à 6.
8. Méthode selon la revendication 8, pour l'identification de la séquence génomique ou de l'ADNc du gène SPG4 de mammifère, notamment de souris.
9. Méthode pour l'identification de mutation portée par le gène SPG4
30 humain, caractérisée en ce qu'elle met en œuvre une séquence nucléique selon l'une des revendications 1 à 6.
10. Méthode selon la revendication 9, pour l'identification de mutation responsable de la paraplégie spastique familiale autosomique dominante.

11. Méthode pour l'identification des séquences d'acide nucléique promotrices et/ou régulatrices de l'expression du gène SPG4, caractérisée en ce qu'elle met en œuvre une séquence nucléique selon l'une des revendications 1 à 6.

12. Acide nucléique identifié par une méthode selon l'une des revendications 8 à 12.

13. Polypeptide codé par un acide nucléique selon l'une des revendications 1 à 3 et 12.

14. Polypeptide selon la revendication 13, caractérisé en ce qu'il comprend une séquence d'acides aminés choisie parmi le groupe comprenant la séquence SEQ ID No. 3, la séquence SEQ ID No. 73, les séquences de polypeptides homologues et variants du polypeptide de séquence SEQ ID No. 3 ou SEQ ID No. 73, et les séquences de leurs fragments d'au moins 10 acides aminés consécutifs.

15. Polypeptide selon la revendication 14, caractérisé en ce qu'il comprend une séquence d'acides aminés choisie parmi le groupe comprenant la séquence SEQ ID No. 3 et la séquence SEQ ID No. 73, lesquelles séquences portant au moins une des mutations dont la nature et la localisation sont identifiées au tableau 5, et les séquences de leurs fragments d'au moins 10 acides aminés consécutifs.

16. Vecteur de clonage et/ou d'expression contenant une séquence d'acide nucléique selon l'une des revendications 1 à 3, et 12.

17. Vecteur selon la revendication 16, caractérisé en ce qu'il comporte les éléments nécessaires à son expression dans une cellule hôte.

18. Cellule hôte transformée par un vecteur selon la revendication 16 ou 17.

19. Mammifère, excepté l'Homme, caractérisé en ce qu'il comprend une cellule selon la revendication 18.

20. Mammifère, excepté l'Homme, selon la revendication 16, comprenant une cellule transformée, caractérisé en ce que la séquence de l'un au moins des deux allèles du gène SPG4 contient une au moins des mutations dont la position et la nature sont identifiées au tableau 5 ou identifiées par une méthode selon la revendication 8 ou 9.

21. Utilisation d'une séquence d'acide nucléique selon l'une des revendications 4, 5 et 12 comme sonde ou amorce, pour la détection et/ou l'amplification de séquences d'acide nucléique.

22. Utilisation d'une séquence d'acide nucléique selon l'une des revendications 1 à 6, et 12, pour le criblage de banque génomique ou d'ADNc.

24. Méthode de production d'un polypeptide recombinant, caractérisée en ce que l'on cultive une cellule transformée selon la revendication 18 dans des conditions permettant l'expression dudit polypeptide recombinant et que l'on récupère ledit polypeptide recombinant.

10 26. Anticorps mono- ou polyclonaux ou leurs fragments, anticorps
chimériques ou immunoconjugués, caractérisés en ce qu'ils sont capables de
reconnaître spécifiquement un polypeptide selon l'une des revendications 13 à 15, et
25

28. Méthode de diagnostic génotypique de la PSF-AD associée au gène SPG4, caractérisée en ce que l'on met en œuvre une séquence d'acide nucléique selon l'une des revendications 1 à 6, et 12.

a) le cas échéant, isolement de l'ADN génomique à partir de l'échantillon biologique à analyser, ou obtention d'ADNc à partir de l'ARN de l'échantillon biologique :

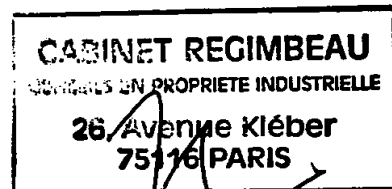
c) analyse des produits d'amplification obtenus et comparaison de leur séquence avec la séquence normale correspondante du gène SPG4.

30 30. Méthode de diagnostic de la PSF-AD associée à une expression
anormale d'un polypeptide codé par le gène SPG4, caractérisée en ce que l'on met n
contact un ou des anticorps selon la revendication 26 avec le matériel biologique à
tester, dans des conditions permettant la formation éventuelle de complexes
immunologiques spécifiques entre ledit polypeptide et l ou lesdits anticorps, et en ce
35 que l'on détecte et/ou quantifie les complexes immunologiques éventuellement formés.

31. Méthode de sélection d'un composé chimique ou biochimique capable d'interagir directement ou indirectement avec un polypeptide selon l'une des revendications 13 à 15, et 25, ou avec un acide nucléique selon l'une des revendications 1 à 6, et 12, et/ou permettant de moduler l'expression ou l'activité de ces polypeptides, caractérisée en ce qu'elle comprend la mise en contact d'une séquence d'acide nucléique selon l'une des revendications 1 à 6, et 12, d'un polypeptide selon l'une des revendications 13 à 15, et 25, d'un vecteur selon l'une des revendications 16 et 17, d'une cellule selon la revendication 18, d'un mammifère selon l'une des revendications 19 et 20 ou d'un anticorps selon la revendication 26, avec un composé candidat et, la détection d'une modification de l'activité dudit polypeptide.

32. Utilisation d'une séquence d'acide nucléique selon l'une des revendications 1 à 6, et 12, d'un polypeptide selon l'une des revendications 13 à 15, et 25, d'un vecteur selon l'une des revendications 16 et 17, d'une cellule selon la revendication 18, d'un mammifère selon l'une des revendications 19 et 20 ou d'un anticorps selon la revendication 26, pour l'étude de l'expression ou de l'activité du gène SPG4.

ORIGINAL



LISTE DE SÉQUENCES

<110> CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE - CNRS

<120> CLONAGE, EXPRESSION ET CARACTÉRISATION DU GÈNE SPG4
RESPONSABLE DE LA FORME LA PLUS FRÉQUENTE DE PARAPLÉGIE
SPASTIQUE AUTOSOMIQUE DOMINANTE

<130> D18374

<160> 105

<170> PatentIn Vers. 2.0

<210> 1
<211> 110000
<212> DNA
<213> Homo sapiens

<220>
<221> exon
<222> (9932)...(10471)

<220>
<221> intron
<222> (10472)...(33718)

<220>
<221> exon
<222> (33719)...(33805)

<220>
<221> intron
<222> (33806)...(35748)

<220>
<221> exon
<222> (35749)...(35832)

<220>
<221> intron
<222> (35833)...(45022)

<220>
<221> exon
<222> (45023)...(45118)

<220>
<221> intron
<222> (45119)...(60863)

<220>
<221> exon
<222> (60864)...(61051)

<220>
<221> intron
<222> (61052)...(61927)

<220>
<223> Amorce

<400> 70
gtgtagcaga tcaacatag

19

<210> 71
<211> 20
<212> ADN
<213> Artificial Sequence

<220>
<223> Amorce

<400> 71
catcttcaag tttggtgcac

20

<210> 72
<211> 1689
<212> ADN
<213> souris

<220>
<223> ADN Spg4 incomplet.

<400> 72
aggccgagag cgtccgcgtc ttccacaagc aggccttcga gtacatctcc attgccctgc 60
gcatcgacga ggaagagaaa gcaggacaga aggaacaagc tgtggaatgg tataagaaag 120
gtatcgaaaga actggaaaaa ggaatcgctg ttatagttac gggccaaggt gaacagtatg 180
aaagagctag acgtcttcaa gccaaaatga tgactaattt agttatggcc aaggaccgtt 240
tacaacttct agagaagctg caaccagttt tgcaattttc caagtcacag acggacgtct 300
ataacgagag tactaacctg acatgccgca atggacatct ccagtcagaa agtggagcag 360
ttccgaagag gaaagacccc ttaacacatg ctagtaattc attgcctcga tcaaaaactg 420
tcctgaaaag tggctccgca gggctctccg gtcaccacag ggcgcctagt tgcagtgggt 480
tgtccatggt ttctggagca agaccgggac ctggtcctgc agctaccaca cataagggta 540
ctccaaaacc aaatagaacc acaaaacctt ctactccac aactgcagtt cggaaaaaga 600
aagacttgaa aaatttttagg aatgtggaca gcaatcttgc taaccttata atgaatgaaa 660
ttgttgacaa tgggacagct gttaagtttg atgacatagc cgggcaggag ctggcaaagc 720
aagcgctgca ggagattgtc atccttcctt ctctgcggcc tgagttgttc acagggctca 780
gagctcctgc tagaggcttg ttactcttcg gtccgccagg aaacggaaaa acaatgctgg 840
ctaaagcagt agctgcagag tctaattgca cttttttcaa cataagtgtt gccagtttaa 900
cttcaaaata tgtgggagaa ggagagaaat tggtagagac tctctttgct gtggctcgag 960
aacttcaacc atctataatt tttatagatg aagttgacag tcttttgtgt gagagacggg 1020
aaggggagca cgacgctagc agacggctaa agacggaatt tttaatagaa tttgacgggg 1080
tgcaatctgc tggagatgac agagtacttg taatgggtgc aactaacagg cccaagagc 1140
ttgatgaagc tgttctcagg cgtttcatta aacgggtata tgtgtcetta ccaaatgagg 1200
agacaagact ccttctgctt aaaaacctgt tgtgtaaaca aggaagtcca ctgaccctaa 1260
aagaactcgc acagcttgct agaattgaccg atggatactc tgggaagtgt ctgaccgctt 1320
tggccaagga tgcagccctg ggtcctatcc gagaactgaa gccagagcag gtgaagaata 1380
tgtctgccag tgagatgaga aatattcgat tatctgactt cacagaatcc ttaaaaaaga 1440
taaaacgcag tgtgagtcct cagaccttag aagcatacat acgctggaac aaggattttg 1500
gagacaccac tgtttaaagg aatggatgcc tctgtgagcc catagaacat cgcacttcac 1560
aggaacaag agctttggct acaggaaccc agacttcgtt tacaggacgt tttagagttt 1620
tcatttttgt gcaccaaact tgaagaggaa caagaagaca gacctaaata aaatatgcaa 1680
tatgaatgg 1689

<210> 73
<211> 504
<212> PRT

<213> souris

<220>

<223> Spastin murine incomplète.

<400> 73

Ala	Glu	Ser	Val	Arg	Val	Phe	His	Lys	Gln	Ala	Phe	Glu	Tyr	Ile	Ser	1	5	10	15
Ile	Ala	Leu	Arg	Ile	Asp	Glu	Glu	Glu	Lys	Ala	Gly	Gln	Lys	Glu	Gln	20	25	30	
Ala	Val	Glu	Trp	Tyr	Lys	Lys	Gly	Ile	Glu	Glu	Leu	Glu	Lys	Gly	Ile	35	40	45	
Ala	Val	Ile	Val	Thr	Gly	Gln	Gly	Glu	Gln	Tyr	Glu	Arg	Ala	Arg	Arg	50	55	60	
Leu	Gln	Ala	Lys	Met	Met	Thr	Asn	Leu	Val	Met	Ala	Lys	Asp	Arg	Leu	65	70	75	80
Gln	Leu	Leu	Glu	Lys	Leu	Gln	Pro	Val	Leu	Gln	Phe	Ser	Lys	Ser	Gln	85	90	95	
Thr	Asp	Val	Tyr	Asn	Glu	Ser	Thr	Asn	Leu	Thr	Cys	Arg	Asn	Gly	His	100	105	110	
Leu	Gln	Ser	Glu	Ser	Gly	Ala	Val	Pro	Lys	Arg	Lys	Asp	Pro	Leu	Thr	115	120	125	
His	Ala	Ser	Asn	Ser	Leu	Pro	Arg	Ser	Lys	Thr	Val	Leu	Lys	Ser	Gly	130	135	140	
Ser	Ala	Gly	Leu	Ser	Gly	His	His	Arg	Ala	Pro	Ser	Cys	Ser	Gly	Leu	145	150	155	160
Ser	Met	Val	Ser	Gly	Ala	Arg	Pro	Gly	Pro	Gly	Pro	Ala	Ala	Thr	Thr	165	170	175	
His	Lys	Gly	Thr	Pro	Lys	Pro	Asn	Arg	Thr	Asn	Lys	Pro	Ser	Thr	Pro	180	185	190	
Thr	Thr	Ala	Val	Arg	Lys	Lys	Lys	Asp	Leu	Lys	Asn	Phe	Arg	Asn	Val	195	200	205	
Asp	Ser	Asn	Leu	Ala	Asn	Leu	Ile	Met	Asn	Glu	Ile	Val	Asp	Asn	Gly	210	215	220	
Thr	Ala	Val	Lys	Phe	Asp	Asp	Ile	Ala	Gly	Gln	Glu	Leu	Ala	Lys	Gln	225	230	235	240
Ala	Leu	Gln	Glu	Ile	Val	Ile	Leu	Pro	Ser	Leu	Arg	Pro	Glu	Leu	Phe	245	250	255	
Thr	Gly	Leu	Arg	Ala	Pro	Ala	Arg	Gly	Leu	Leu	Leu	Phe	Gly	Pro	Pro	260	265	270	
Gly	Asn	Gly	Lys	Thr	Met	Leu	Ala	Lys	Ala	Val	Ala	Ala	Glu	Ser	Asn	275	280	285	

Ala Thr Phe Phe Asn Ile Ser Ala Ala Ser Leu Thr Ser Lys Tyr Val
 290 295 300
 Gly Glu Gly Glu Lys Leu Val Arg Ala Leu Phe Ala Val Ala Arg Glu
 305 310 315 320
 Leu Gln Pro Ser Ile Ile Phe Ile Asp Glu Val Asp Ser Leu Leu Cys
 325 330 335
 Glu Arg Arg Glu Gly Glu His Asp Ala Ser Arg Arg Leu Lys Thr Glu
 340 345 350
 Phe Leu Ile Glu Phe Asp Gly Val Gln Ser Ala Gly Asp Asp Arg Val
 355 360 365
 Leu Val Met Gly Ala Thr Asn Arg Pro Gln Glu Leu Asp Glu Ala Val
 370 375 380
 Leu Arg Arg Phe Ile Lys Arg Val Tyr Val Ser Leu Pro Asn Glu Glu
 385 390 395 400
 Thr Arg Leu Leu Leu Leu Lys Asn Leu Leu Cys Lys Gln Gly Ser Pro
 405 410 415
 Leu Thr Gln Lys Glu Leu Ala Gln Leu Ala Arg Met Thr Asp Gly Tyr
 420 425 430
 Ser Gly Ser Asp Leu Thr Ala Leu Ala Lys Asp Ala Ala Leu Gly Pro
 435 440 445
 Ile Arg Glu Leu Lys Pro Glu Gln Val Lys Asn Met Ser Ala Ser Glu
 450 455 460
 Met Arg Asn Ile Arg Leu Ser Asp Phe Thr Glu Ser Leu Lys Lys Ile
 465 470 475 480
 Lys Arg Ser Val Ser Pro Gln Thr Leu Glu Ala Tyr Ile Arg Trp Asn
 485 490 495
 Lys Asp Phe Gly Asp Thr Thr Val
 500

<210> 74

<211> 24

<212> ADN

<213> Homo sapiens

<220>

<223> Site accepteur d'épissage du gène SPG4.

<400> 74

attttttatt ttaaagcagg acag

24

<210> 75

<211> 24

<212> ADN

<213> Homo sapiens

<220>
<223> Site accepteur d'épissage du gène SPG4.

<400> 75
aatttttttc tttcaggtga acag 24

<210> 76
<211> 24
<212> ADN
<213> Homo sapiens

<220>
<223> Site accepteur d'épissage du gène SPG4.

<400> 76
cttctctgtt gcatagagaa gatg 24

<210> 77
<211> 24
<212> ADN
<213> Homo sapiens

<220>
<223> Site accepteur d'épissage du gène SPG4.

<400> 77
actttttcct tgtcagaaag tgga 24

<210> 78
<211> 24
<212> ADN
<213> Homo sapiens

<220>
<223> Site accepteur d'épissage du gène SPG4.

<400> 78
ttttgtatcc tttaagggtg ctcc 24

<210> 79
<211> 24
<212> ADN
<213> Homo sapiens

<220>
<223> Site accepteur d'épissage du gène SPG4.

<400> 79
agggtcttgtt tcttagtgga acag 24

<210> 80
<211> 24
<212> ADN
<213> Homo sapiens

<220>
<223> Site accepteur d'épissage du gène SPG4.

<400> 80
agtatatatt ttttagttgt tcac 24

<210> 81
<211> 24
<212> ADN
<213> Homo sapiens

<220>
<223> Site accepteur d'épissage du gène SPG4.

<400> 81
cttgtgattt ttaaaggcta aagc 24

<210> 82
<211> 24
<212> ADN
<213> Homo sapiens

<220>
<223> Site accepteur d'épissage du gène SPG4.

<400> 82
taatgctttg ttttaggtgg gaga 24

<210> 83
<211> 24
<212> ADN
<213> Homo sapiens

<220>
<223> Site accepteur d'épissage du gène SPG4.

<400> 83
cttgtatttc ctctagatga agtt 24

<210> 84
<211> 24
<212> ADN
<213> Homo sapiens

<220>
<223> Site accepteur d'épissage du gène SPG4.

<400> 84
gattttttgc ttgtaggtac agtc 24

<210> 85
<211> 24
<212> ADN
<213> Homo sapiens

<220>
<223> Site accepteur d'épissage du gène SPG4.

<400> 85
ggattttttt ttttaggcgt ttca 24

<210> 86
<211> 24
<212> ADN
<213> Homo sapiens

<220>
<223> Site accepteur d'épissage du gène SPG4.

<400> 86
ttttaaatatt tttcagacaa gact 24

<210> 87
<211> 24
<212> ADN
<213> Homo sapiens

<220>
<223> Site accepteur d'épissage du gène SPG4.

<400> 87
tccttccctt cctcagaatg actg 24

<210> 88
<211> 24
<212> ADN
<213> Homo sapiens

<220>
<223> Site accepteur d'épissage du gène SPG4.

<400> 88
cttttatggt ttacagaact aaaa 24

<210> 89
<211> 24
<212> ADN
<213> Homo sapiens

<220>
<223> Site accepteur d'épissage du gène SPG4.

<400> 89
cttttttaaaa atctagatga gaaa 24

<210> 90
<211> 24
<212> ADN
<213> Homo sapiens

<220>
<223> Site donneur d'épissage du gène SPG4.

<400> 90
tgagaaaggt aactaggggg ctgg 24

<210> 91
<211> 24
<212> ADN
<213> Homo sapiens

<220>
<223> Site donneur d'épissage du gène SPG4.

<400> 91
aggacaaggt aagattgtat ttgt 24

<210> 92
<211> 24
<212> ADN
<213> Homo sapiens

<220>
<223> Site donneur d'épissage du gène SPG4.

<400> 92
acttctaggt atcaattaat gtat 24

<210> 93
<211> 24
<212> ADN
<213> Homo sapiens

<220>
<223> Site donneur d'épissage du gène SPG4.

<400> 93
ccagtcaggt gggtttaggt taac 24

<210> 94
<211> 24
<212> ADN
<213> Homo sapiens

<220>
<223> Site donneur d'épissage du gène SPG4.

<400> 94
ctcataaggt attctgggac agta 24

<210> 95
<211> 24
<212> ADN
<213> Homo sapiens

<220>
<223> Site donneur d'épissage du gène SPG4.

<400> 95
gtggacaagt aagttttgcc atct 24

<210> 96
<211> 24
<212> ADN
<213> Homo sapiens

<220>
<223> Site donneur d'épissage du gène SPG4.

<400> 96
ggcctgaggt aagaacttta tatt

24

<210> 97
<211> 24
<212> ADN
<213> Homo sapiens

<220>
<223> Site donneur d'épissage du gène SPG4.

<400> 97
caatgctggt aagggttctc ttca

24

<210> 98
<211> 24
<212> ADN
<213> Homo sapiens

<220>
<223> Site donneur d'épissage du gène SPG4.

<400> 98
caaaatacgt gagtgctctg ttcc

24

<210> 99
<211> 24
<212> ADN
<213> Homo sapiens

<220>
<223> Site donneur d'épissage du gène SPG4.

<400> 99
ttttataggt aagaacatat ttcc

24

<210> 100
<211> 24
<212> ADN
<213> Homo sapiens

<220>
<223> Site donneur d'épissage du gène SPG4.

<400> 100
ttgatggtgt aagtgttgat tatg

24

<210> 101
<211> 24
<212> ADN
<213> Homo sapiens

<220>
<223> Site donneur d'épissage du gène SPG4.

<400> 101
gttctcaggt agggagattt atat

24

<210> 102
<211> 24
<212> ADN
<213> Homo sapiens

<220>
<223> Site donneur d'épissage du gène SPG4.

<400> 102
atgaggaggt atgtatctgt gttt 24

<210> 103
<211> 24
<212> ADN
<213> Homo sapiens

<220>
<223> Site donneur d'épissage du gène SPG4.

<400> 103
cttgctaggt gagtaatttg gatt 24

<210> 104
<211> 24
<212> ADN
<213> Homo sapiens

<220>
<223> Site donneur d'épissage du gène SPG4.

<400> 104
tatccgaggt aggtatacaa gagc 24

<210> 105
<211> 24
<212> ADN
<213> Homo sapiens

<220>
<223> Site donneur d'épissage du gène SPG4.

<400> 105
ccagtgaggt atagtatttt acaa 24

REVENDEICATIONS

1. Acide nucléique purifié ou isolé du gène SPG4, caractérisé en ce qu'il comprend au moins 15 nucléotides consécutifs d'une séquence choisie parmi le
5 groupe comprenant la séquence SEQ ID No. 1, les séquences nucléiques homologues ou variantes de l'acide nucléique de séquence SEQ ID No. 1, leur séquence complémentaire et la séquence de leur ARN correspondant.
2. Acide nucléique purifié ou isolé selon la revendication 1, caractérisé en ce qu'il comprend une séquence choisie parmi le groupe comprenant la séquence SEQ
10 ID No. 1, la séquence SEQ ID No. 2, la séquence SEQ ID No. 72, les séquences nucléiques homologues ou variantes des séquences SEQ ID No. 1, SEQ ID No. 2 ou SEQ ID No. 72, leur séquence complémentaire et la séquence de leur ARN correspondant.
3. Acide nucléique purifié ou isolé selon la revendication 1 ou 2,
15 caractérisé en ce qu'il comprend une mutation correspondant à un polymorphisme naturel chez l'Homme.
4. Sonde ou amorce, caractérisée en ce qu'elle comprend une séquence d'un acide nucléique selon l'une des revendications 1 à 3.
5. Sonde ou amorce selon la revendication 4, caractérisée en ce que sa
20 séquence est choisie parmi les séquences SEQ ID No. 4 à SEQ ID No. 71.
6. Site accepteur ou donneur d'épissage, caractérisé en ce que sa séquence est choisie parmi les séquences SEQ ID No. 74 à SEQ ID No. 105.
7. Méthode pour le criblage de banques d'ADNc ou d'ADN génomique ou pour le clonage d'ADNc ou génomique isolé codant pour la spastin, caractérisée en ce
25 qu'elle met en œuvre une séquence nucléique selon l'une des revendications 1 à 6.
8. Méthode selon la revendication 7, pour l'identification de la séquence génomique ou de l'ADNc du gène SPG4 de mammifère, notamment de souris.
9. Méthode pour l'identification de mutation portée par le gène SPG4 humain, caractérisée en ce qu'elle met en œuvre une séquence nucléique selon l'une
30 des revendications 1 à 6.
10. Méthode selon la revendication 9, pour l'identification de mutation responsable de la paraplégie spastique familiale autosomique dominante.
11. Méthode pour l'identification des séquences d'acide nucléique promotrices et/ou régulatrices de l'expression du gène SPG4, caractérisée en ce
35 qu'elle met en œuvre une séquence nucléique selon l'une des revendications 1 à 6.

12. Acide nucléique identifié par une méthode selon l'une des revendications 8 à 11.

13. Polypeptide codé par un acide nucléique selon l'une des revendications 1 à 3 et 12.

5 14. Polypeptide selon la revendication 13, caractérisé en ce qu'il comprend une séquence d'acides aminés choisie parmi le groupe comprenant la séquence SEQ ID No. 3, la séquence SEQ ID No. 73, les séquences de polypeptides homologues et variants du polypeptide de séquence SEQ ID No. 3 ou SEQ ID No. 73, et les séquences de leurs fragments d'au moins 10 acides aminés consécutifs.

10 15. Polypeptide selon la revendication 14, caractérisé en ce qu'il comprend une séquence d'acides aminés choisie parmi le groupe comprenant la séquence SEQ ID No. 3 et la séquence SEQ ID No. 73, lesquelles séquences portant au moins une des mutations correspondant à un polymorphisme naturel chez l'Homme, et les séquences de leurs fragments d'au moins 10 acides aminés consécutifs.

15 16. Vecteur de clonage et/ou d'expression contenant une séquence d'acide nucléique selon l'une des revendications 1 à 3, et 12.

17. Vecteur selon la revendication 16, caractérisé en ce qu'il comporte les éléments nécessaires à son expression dans une cellule hôte.

18. Cellule hôte transformée par un vecteur selon la revendication 16 ou 17.

20 19. Mammifère, excepté l'Homme, caractérisé en ce qu'il comprend une cellule selon la revendication 18.

20. Mammifère, excepté l'Homme, selon la revendication 19, comprenant une cellule transformée, caractérisé en ce que la séquence de l'un au moins des deux allèles du gène SPG4 contient une au moins des mutations correspondant à un polymorphisme naturel chez l'Homme ou identifiées par une méthode selon la revendication 9 ou 10.

25 21. Utilisation d'une séquence d'acide nucléique selon l'une des revendications 4, 5 et 12 comme sonde ou amorce, pour la détection et/ou l'amplification de séquences d'acide nucléique.

30 22. Utilisation d'une séquence d'acide nucléique selon l'un des revendications 1 à 6, et 12, pour le criblage de banque génomique ou d'ADNc.

23. Utilisation d'une séquence d'acide nucléique selon l'une des revendications 1 à 3, et 12, pour la production d'un polypeptide recombinant ou synthétique.

24. Méthode de production d'un polypeptide recombinant, caractérisée en ce que l'on cultive une cellule transformée selon la revendication 18 dans des conditions permettant l'expression dudit polypeptide recombinant et que l'on récupère ledit polypeptide recombinant.

5 25. Polypeptide, caractérisé en ce qu'il est obtenu par une méthode selon la revendication 24.

26. Anticorps mono- ou polyclonaux ou leurs fragments, anticorps chimériques ou immunoconjugués, caractérisés en ce qu'ils sont capables de reconnaître spécifiquement un polypeptide selon l'une des revendications 13 à 15, et
10 25.

27. Méthode pour la détection et/ou la purification d'un polypeptide selon l'une des revendications 13 à 15, et 25, caractérisée en ce qu'elle met en oeuvre un anticorps selon la revendication 26.

28. Méthode de diagnostic génotypique de la PSF-AD associée au gène
15 SPG4, caractérisée en ce que l'on met en oeuvre une séquence d'acide nucléique selon l'une des revendications 1 à 6, et 12.

29. Méthode de diagnostic génotypique de PSF-AD associée à la présence d'au moins une mutation sur une séquence du gène SPG4 à partir d'un prélèvement biologique d'un patient, caractérisée en ce qu'elle comporte les étapes suivantes :
20 a) le cas échéant, isolement de l'ADN génomique à partir de l'échantillon biologique à analyser, ou obtention d'ADNc à partir de l'ARN de l'échantillon biologique ;
b) amplification spécifique de ladite séquence d'ADN du gène SPG4 susceptible de contenir une mutation à l'aide d'amorces selon l'une des revendications 4 et 5 ou d'un acide nucléique selon la revendication 12 ;
25 c) analyse des produits d'amplification obtenus et comparaison de leur séquence avec la séquence normale correspondante du gène SPG4.

30. Méthode de diagnostic de la PSF-AD associée à une expression anormale d'un polypeptide codé par le gène SPG4, caractérisée en ce que l'on met en contact un ou des anticorps selon la revendication 26 avec le matériel biologique à
30 tester, dans des conditions permettant la formation éventuelle de complexes immunologiques spécifiques entre ledit polypeptide et le ou lesdits anticorps, et en ce que l'on détecte et/ou quantifie les complexes immunologiques éventuellement formés.

31. Méthode de sélection d'un composé chimique ou biochimique capable d'interagir directement ou indirectement avec un polypeptide selon l'une des
35 revendications 13 à 15, et 25, ou avec un acide nucléique selon l'une des

revendications 1 à 6, et 12, et/ou permettant de moduler l'expression ou l'activité de ces polypeptides, caractérisée en ce qu'elle comprend la mise en contact d'une séquence d'acide nucléique selon l'une des revendications 1 à 6, et 12, d'un polypeptide selon l'une des revendications 13 à 15, et 25, d'un vecteur selon l'une des
5 revendications 16 et 17, d'une cellule selon la revendication 18, d'un mammifère selon l'une des revendications 19 et 20 ou d'un anticorps selon la revendication 26, avec un composé candidat et, la détection d'une modification de l'activité dudit polypeptide.

32. Utilisation d'une séquence d'acide nucléique selon l'une des revendications 1 à 6, et 12, d'un polypeptide selon l'une des revendications 13 à 15, et
10 25, d'un vecteur selon l'une des revendications 16 et 17, d'une cellule selon la revendication 18, d'un mammifère selon l'une des revendications 19 et 20 ou d'un anticorps selon la revendication 26, pour l'étude de l'expression ou de l'activité du gène SPG4.

REVENDEICATIONS

1. Acide nucléique purifié ou isolé du gène SPG4, caractérisé en ce qu'il comprend au moins 15 nucléotides consécutifs d'une séquence choisie parmi le
5 groupe comprenant la séquence SEQ ID No. 1, les séquences nucléiques homologues ou variantes de l'acide nucléique de séquence SEQ ID No. 1, leur séquence complémentaire et la séquence de leur ARN correspondant.

2. Acide nucléique purifié ou isolé selon la revendication 1, caractérisé en ce qu'il comprend une séquence choisie parmi le groupe comprenant la séquence SEQ
10 ID No. 1, la séquence SEQ ID No. 2, la séquence SEQ ID No. 72, les séquences nucléiques homologues ou variantes des séquences SEQ ID No. 1, SEQ ID No. 2 ou SEQ ID No. 72, leur séquence complémentaire et la séquence de leur ARN correspondant.

3. Acide nucléique purifié ou isolé selon la revendication 1 ou 2,
15 caractérisé en ce qu'il comprend une mutation correspondant à un polymorphisme naturel chez l'Homme.

4. Sonde ou amorce, caractérisée en ce qu'elle comprend une séquence d'un acide nucléique selon l'une des revendications 1 à 3.

5. Sonde ou amorce selon la revendication 4, caractérisée en ce que sa
20 séquence est choisie parmi les séquences SEQ ID No. 4 à SEQ ID No. 71.

6. Site accepteur ou donneur d'épissage, caractérisé en ce qu'il comprend une séquence d'un acide nucléique selon la revendication 1 choisie parmi les séquences SEQ ID No. 74 à SEQ ID No. 105.

7. Méthode pour le criblage de banques d'ADNc ou d'ADN génomique ou
25 pour le clonage d'ADNc ou génomique isolé codant pour la spastin, caractérisée en ce qu'elle met en œuvre une séquence nucléique selon l'une des revendications 1 à 6.

8. Méthode selon la revendication 7, pour l'identification de la séquence génomique ou de l'ADNc du gène SPG4 de mammifère, notamment de souris.

9. Méthode pour l'identification de mutation portée par le gène SPG4
30 humain, caractérisée en ce qu'elle met en œuvre une séquence nucléique selon l'une des revendications 1 à 6.

10. Méthode selon la revendication 9, pour l'identification de mutation responsable de la paraplégie spastique familiale autosomique dominante.

11. Méthode pour l'identification des séquences d'acide nucléique
35 promotrices et/ou régulatrices de l'expression du gène SPG4, caractérisée en ce qu'elle met en œuvre une séquence nucléique selon l'une des revendications 1 à 6.

FIGURE 1A

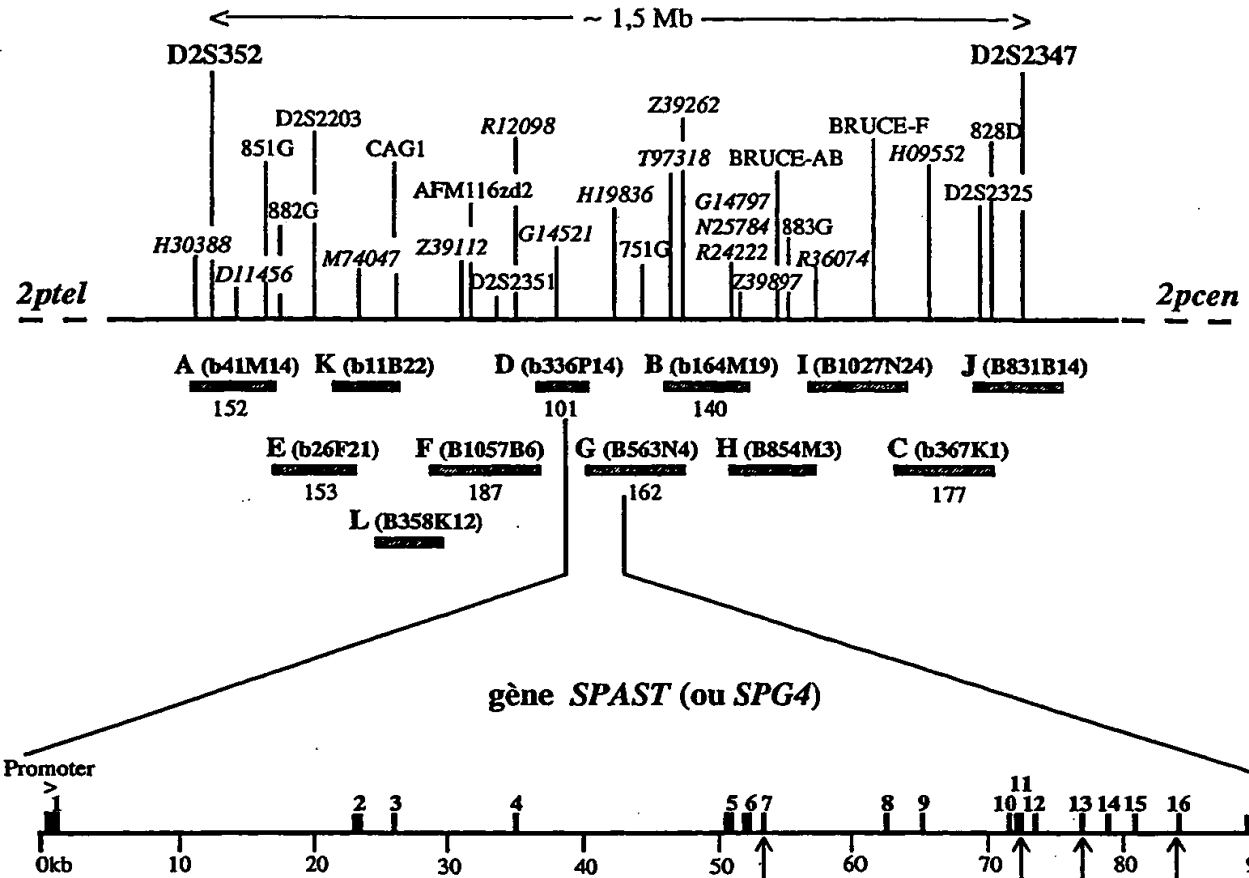
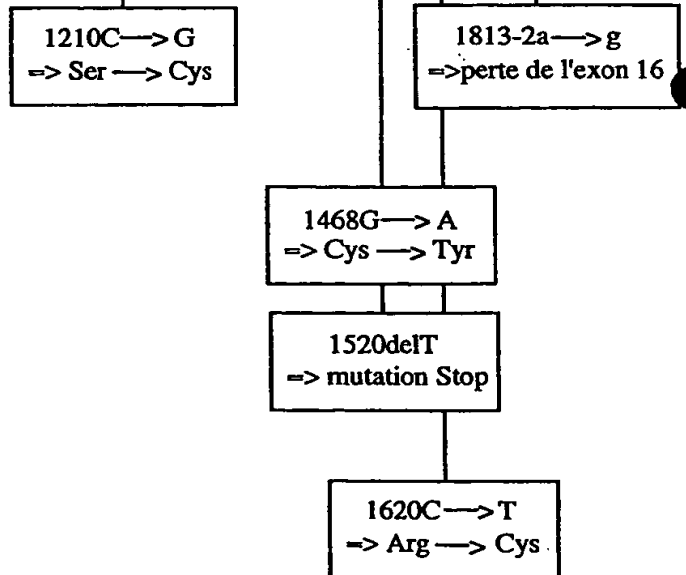


FIGURE 1B

FIGURE 1C



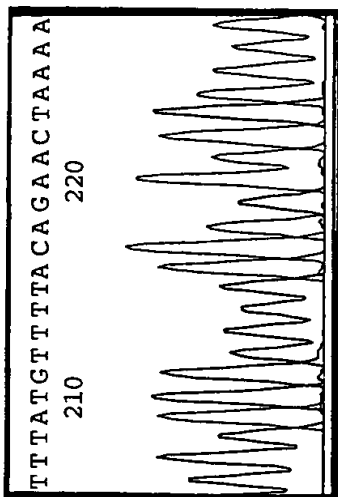
1 GCTCCTGAGACCGGCGGGCACACGGGGTCTGTGGCCCCCGCGTAGCAGTGGCTGCCGCGCTGCTTGGTTCCCGTCCGCTCTGCGGGAGGCGGG 95
 1 TTATGGCGGCGGCGGCGAGTGTGAATGAATTCTCCGGTGGACGAGGGAAGAAAGGCTCCGCGCGGCCAGCAACCGGCTGCTCC 190
 1 M N S P G G R G K K K G S G G A S N P V P P
 23 R P P P C L A P A P P A A G P A P P P E S P H K R N L Y Y F S 285
 CTTACCGCTGTTTGTAGGCTTCGCGCTGCTGCGTTTGGTGGCTTCCACCTGGGGCTCTCTCTGTGGTCTGCGAGCGCTTCTCCGCGGCC 380
 55 Y P L F V G F A L L R L V A F H L G L L F V W L C Q R F S R A
 CTCATGGCAGCCAAGAGGAGCTCCGGGGCGCGCCAGCAGCTGCTCGGGCTCGGGCCCGCGCGCGGTGCCGGGGCGGAGGCGGAGCGCGTCCG 475
 86 L M A A K R S S G A A P A P A S A S A P A P V P G G E A E R V R
 AGTCTTCCACAAACAGGCTTCGAGTATCATCTCCATGCGCTCGAGCATGAGGATGAGAAAGAGGACAGAGGCAAGCTGTGGAATGGT 570
 118 V F H K Q A F E Y I S I A L R I D E D E K A 2 G Q K E Q A V E W Y
 ATAAGAAAGGTATTGAAGAACTGGAAGAAAGGAATAGCTGTATAGTTACAGGACAAGGTGAACAGTGTGAAGAGCTAGACGCTTCAAGCTAAA 665
 150 K K G I E E L E K G I A V I V T G Q G 3 E Q C E R A R R L Q A K
 ATGAGTACTAATTTGGTTATGGCCAAAGGACCGCTTACAACITCTAGAGAAAGTGGACAGTCTTCCCAAGTCACAAACGGAAGCTCTA 760
 181 M M T N L V M A K D R L Q L L E K M Q P V L P F S K S Q T D V Y
 TAATGACAGTACTAAGTGGCATGCCGCAATGGACATCTCCAGTCAAGAGTGGAGCTGTTCACAAAAGAAAAGACCCCTTAACACACACTAGTA 855
 213 N D S T N L A C R N G H L Q S E 5 S G A V P K R K D P L T H T S N
 ATTCACTGCCCTCGTTTCAAAAACAGTTATGAAAACAGTGGATCGAGCCCTTTACGCCACCATAGAGCAGCTAGTGGTGGTTATCCATGGTT 950
 245 S L P R S K T V M K T G S A G L S G H H R A P S Y S G L S M V
 TCTGGAGTGAACAGGATCTGGTCTGCTCTACCACTCATAAGTGTACTCCGAAAACAAATAGGACAAATAACCTTCTACCCCTACAACTGC 1045
 276 S G V K Q G S G P A P T T H K G 6 T P K T N R T N K P S T P T T A
 TACTCTGAAGAAAAGAGCTTGAAGAATTTAGGAATGTGGACAGCACTTGAACCTTATAATGAATGAATGGTGGAGCTTGAAGCTG 1140
 308 T R K K K D L K N F R N V D S N L A N L I M N E I V D N G 7 T A V
 TTAATTTGATGATATAGCTGGTCAAGACTTGGCAAAACAGCATGTCAAGAAATGTTATTCTCTCTGAGGCTGAGTTGTTACAGGG 1235
 340 K F D D I A G Q D L A K Q A L Q E I V I L P S L R P E L S F T G
 CTTAGAGCTCCTGCCAGGGCTGTACTCTTTGGTCCACCTGGGAATGGGAAGCAATGTCTGCTAAGCAGTAGCTGCAGAACTCGAATGCAAC 1330
 371 L R A P A R G L L L F G P P G N G K T M L A 9 K A V A A E S N A T
 CTTCTTAATATAAGTGTGCAAGTTAACTTCAAAATAGTGGGAGAAGGAGAGAAATGGTGGGGCTCTTTTGTGTGGCTCGAGAACTTC 1425
 403 F F N I S A A S L T S K Y V 10 G E G E K L V R A L F A V A R E L Q
 AACCTTCTATAATTTTATAGTGAAGTTGATAGCTTTTGTGTGAAGAAAGAGGGGACGATGCTAGTAGAGCCTAAAACTGAATTT 1520
 435 P S I I F I D I E V D S L C E R R E G E H D A S R R L K T E F
 CTAATAGAATTTGATGGTTTACAGTCTGCTGGAGATGACAGACTTGTAAATGGGTGCAACTAATAGGCCACAAGAGCTTGTAGAGGCTGTTCT 1615
 466 L I E F D G V 12 Q S A G D D R V L V M G A T N R P Q E L D E A V L
 CAGCGTTTCACTCAAAACGGGTATATGTCTTTACCAATGAGGAGTCAAGACTTTTGTCTTAAAAATCTGTTATTAACAAGGAAGTCCAT 1710
 498 R 13 R F I K R V Y V S L P N E E T 14 R L L L K N L L C K Q G S P L
 TGACCCAAAAGAACTAGCACAACTTGCTAGTACTGATGATGATACTCAGGAAGTGACCTAACAGCTTTGGCAAAAGATGCAGCACTGGGTCT 1805
 530 T Q K E L A Q L A R M 15 T D G Y S G S D L T A L A K D A A L G P
 ATCCGAGTAACTAAAACAGAAACAGGTGAAGAAATATGTCTGCCAGTGAATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAAGAAAT 1900
 561 I R E 16 L K P E Q V K N M S A S E M 17 R N I R L S D F T E S L K K I
 AAAACGAGCGCTCAGCCCTCAAACCTTTAGAAGCGTACATAGTGGAAACAGGACTTTGGAGATACCACTGTTTAAGGAAATACCTTTGTAAACC 1995
 593 K R S V S P Q T L E A Y I R W N K D F G D T T V *
 TGCAGAACATTTTACTTAAAGAGGAAACACAAGATCTTCAATGAACGTATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTATAGTCT 2090
 TACATATTTGTGCACCAAACTTGAAGATGAACAGAAACAGACTTAAACAAAATATACAATGCAAAATGTAATTTTGTGTGTTAAGCCCTTGC 2185
 CTTGATGGTCAAGTTATCCCAATGGACACTAAGTTAGAGCACAACAAAACCTGATTCTGGTCTCTTTTACCAATATAATCATAAATGTAAATAAT 2280
 AATTTGTATATTTGTGTCAGATGAAGTATTCCAGGAACAGTGAATGGTAGAAGACACAAGAACATTGTTTGTGTTCTCTGATGTTTTTTC 2375
 TTAATAAGTAATTTCTCTACTTTTCTTTCTACTGTTGCTTAACTACAGGTGATTGGAATGCCAAACACTCTTAAGTTTATTTTCTTTTTC 2470
 GTTTTATAATTCAGTGTGTCGCAAAATGAAACTTTTCTCAAGTAACTGTAATAGGAAAAGTTATTTTGTAGAGTTCTTCTTCAATAATCTACA 2565
 GACATTAAACAATTTGTTGTTCTTTTACCTTTTATTTTCTATTACCTTGCTACCAACAGCTTTAGATAGCAATATAATAGCAAAAAGCAAA 2660
 TATGGTAAAAATAGAGAAGGTTTGAAGGTTTGAAGTTTACTCTGTCTATATAACATGTAGATCAGTCTTCAATGTGACCTGCAGTATTTTCTTAAAT 2755
 GTATTGTGCAAGAACTGTTGTAGACTGTTAACTTCTCTCTGATGGAATTTATTTCTGCAAGAAATTTCTGATATTTAAGAGAGCCAAATTTTA 2850
 ACTGCTGTGAAAATGTTTCCAGTGAAGAGAAAGGAAATACTAGGAACTAAGACATTCTAATTTATGCTTATTACTTTCTAATTTTACAGGA 2945
 TAATTATAAGCAAGTGAACCTACCATCTTTTATTTCTTAATAATTATTAATCCCTTCAATGAACTTTAAAAAACTGAATTTTATACATGGCAT 3040
 ACATTTTCTAGTTCTCTGCTTTTATTAACCTCAAAAGTTCTAGTTCTAGTCTGTTGATCTGCTTTTGTCTCCCAAAATGTACAGTAAT 3135
 TCCATTTGTTGTATAAATAAGCTGGATTTTCATTATAAAAAATGTCATTGTAGGGAGTAGAGACTCATATCATGGCCTTTTAAATATTGTAATA 3230
 AAGGCAATAGATATTTGCCCTTAGTTTACTGG 3263

FIGURE 2

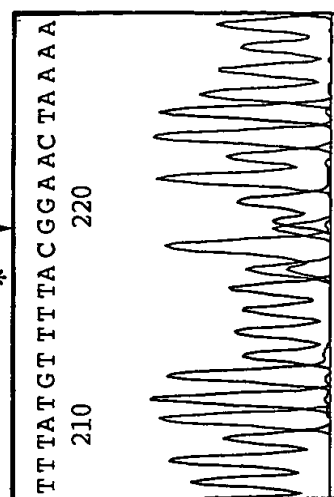
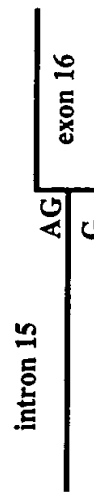
CABINET REGIMBEAU
 ORIGINAL

B

Site accepteur d'épissage normal



Site accepteur d'épissage muté



A

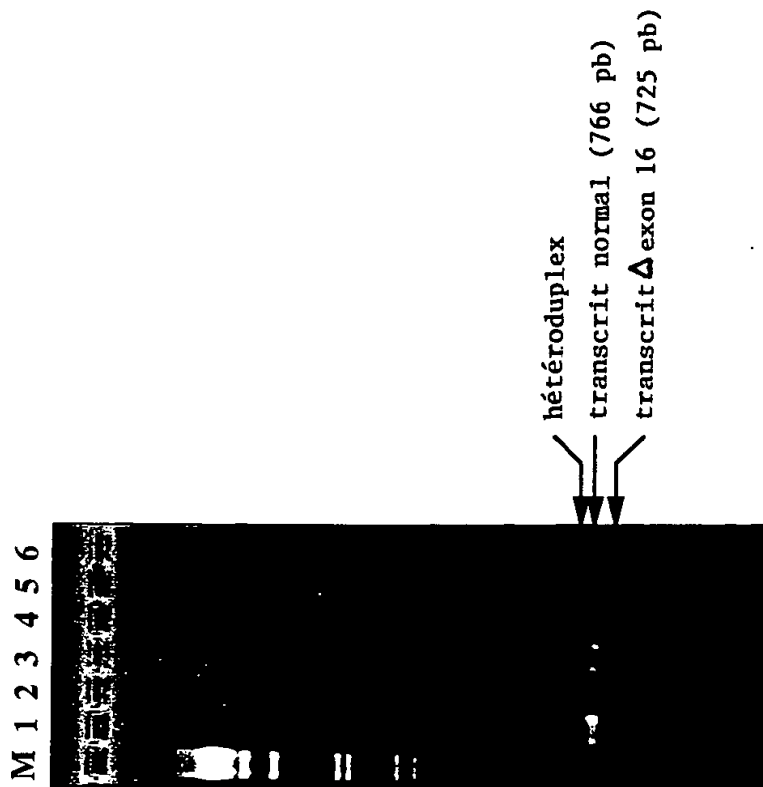


FIGURE 3

[illegible]

FIGURE 4A

```

AFG3_YEAST ---MMHWQRYARGAPR-----SLTSL-----PGKASRISTVKP-----VLESMPVHQRILQTLGSLATR-----HTIHR--S
RCAL_YEAST -MLLLSWSRIATKVVRFPVFRSYYGLTHIKSLRTQYRLNRLQENKSGHKNEDHNEADKLHKEIPTDEEVEAIRKQVEKYIEQTKNNTIPARWK
PARAPLEGIN MAVLLLLLRLALRERGPGGPR-----PLWGPGPAMS--PGFPARPGRGRP-----YMASRPPGD--LAEAGGRALQ-----S
YME1_YEAST ---MNVSKILVSPVTVTN-----VLRFAPRLFPQIGASLLVQKKWA-----LESKFFYRFYSEKNSGEMPP-----K
SPAST_HUMAN ---MNSPGGRGKKGSGGASN-----FVFRPPFPCLAPAPPAAGPAPP-----ESPHKNLYYFSPYFLFVGALLRLVAFHLGLLVWLC

AFG3_YEAST TQIRSFHISWRLN-----ENRPNKE--GEGKXNGNKDNNSNKEDGKDKRNEFGSLS-----EYFRSKEFANTMFLTIGFTIIF
RCAL_YEAST EQKRKIDESIRLEDAVLKQESNRHQERKEKEEENGPSKAKSNRTKEQGYFEGNHSRNIIPFPFPFPKPLNDPSGNPVSKNVNLFQIGLTLTFL
PARAPLEGIN LQLRLLTPTTFEGINGLLKQH--LVQNFVRLWQLGGTFYFNTSRHKQKHKEKDKSKGKAPFEDE-----EERRRERDDQMIRBELRLTLV
YME1_YEAST KEADSSGKASNKST-----ISSIDNSQF--PPPSNTNDKTKQANVAVSHAMLATREGEA-----NKDLTSPDAQAAFYKLLQSNY
SPAST_HUMAN QRFSRALMAAKRSSGAAPAP--ASASAPAPVP--GGEARVRVPHKQAFEYISIALRIDEDEKAG-----KQEQAWEWYKKGIELEKGIAY

AFG3_YEAST TLLTPSSNNSGDDSNELVTQDFKTKYLEKGLVSKIYV--NKF--LVEAELVH-----TKQVVSFTIGSV----DIFREQ--MDQIQDLLNI
RCAL_YEAST LSFLDLLNSLEEQS-EITWQDFREKLLAKGYVAKLIV--NKS--MVKVMLNDNGKNQADNYGRNFYFTIGSI----DSFEHK--LQKAQDELDI
PARAPLEGIN IAVVMGSLNALSTSGGSIWNDFVHEMLAKGEVQVQV--VESDVVEVYLHPGAVVFGPRRLALMYRMQVANI-----DKFEK--LRAADELNI
YME1_YEAST PQYVVSSEFTFGIASSPECMELYMALQRIGRSEADA--EQN--LDTASSAG-----AVNPSLASSSSNQ----SGYEGN--FFSNYSPLYG
SPAST_HUMAN IVTGQGEQCEARERLQAKMMTNLVMAKRLQLLEKMQP--LPFSKSQTDVYNDSTN--LACRNGHLOSEGAVFPRKRDPLTHTSBSLPSKTVMKT

AFG3_YEAST PPRDRIPIKYIERSSPFTFLFPF---LPTIILLGGLYFITERKINSSPPNANGGGGGGLGGMFNVGKSRALKFKETDIK--IS--KNVA--CDE
RCAL_YEAST DKDFRIPVLYVQEGNWAAMFQI---LPTVLMIAGIWLTERS-----AQAGGSGRGGIFGLSRKAKKFNTEDVK--IK--KDAV--CDE
PARAPLEGIN EAKDRIPVSYKRTGFFGNALYSV---GMTAVGLAILWVVPFLAG-----MTGREGCSAFHQLKMARFTIVDGKMGK--VS--KDAV--MHE
YME1_YEAST SRKE--PLHVVSSESTFTVVSBN---VKWLLVFGILTYSFSEC-----PKYITENTLLKSSEVADKSDVDVAKTN--VK--DVCV--CDE
SPAST_HUMAN GSAGLSCGHRAPSYSGLSMVSGVKQSGPAPTTHKGTPTNRTNKPSTPTTATRKKDLKHFNVDSHNLANLIMNEIVDNGTAVK--DIA--QDL

AFG3_YEAST KQEIIMF--HFLKN--GKYTKLGAKIPRAIS--T--L--K--T--G--ANVP--LSVSGSEFVEMFV--V--ASEVD--TO--SMA--S--IF--I
RCAL_YEAST KEEIMF--SFLKE--SRIEKGAKIPRAIS--T--L--K--T--G--AGVP--YFVSGSEFVEMFV--V--AAEVD--KT--ENA--S--VF--I
PARAPLEGIN KLEVR--P--DYLS--ERFLQLGAKVPRAL--L--C--L--K--V--T--AQVP--LAMAGPEFVEVIG--L--AAEVS--KE--ABA--C--VY--I
YME1_YEAST BAELEI--DFLKD--TKYESLGGKLPK--VL--T--Y--L--E--T--G--AGVD--PFMSGSEFDEVIV--V--AKEID--AQ--SRA--A--IF--L
SPAST_HUMAN KQALO--I--ILPSLE--ELFTGLRAPA--R--LL--F--N--M--K--V--A--SNAT--FNISAASLTSKYV--E--EKLVA--AV--ELO--S--IF--V

AFG3_YEAST AIGKE--GKGGALGGAANDEREATLNQL--V--M--FTTS--DQVVVLAG--EPDV--M--LM--PG--FDRHIQIDS--DVNG--QOIYVLKRLNLDPLLT
RCAL_YEAST AIGKA--QKG--NFSGANDERENTLQNM--V--M--FTPA--DEVVVLAG--EPDI--K--LL--PG--FDRHINIDK--ELEG--KAIFAVHLHLKLAGEIF
PARAPLEGIN AVGKK--STT--MSGFSNTEERETLNQL--V--M--MGTT--DEVIVLAS--RADI--G--LM--PG--LDRHVVIDL--TLQE--REIFEQHLKSLKLTQSS
YME1_YEAST AIGCK--NPK-----DQAYAKOTLNQL--V--L--FSQT--SGIIIGAG--FPEA--K--LT--PG--FDKVVNVDL--DVRG--ADILKHHMKKITLADNV
SPAST_HUMAN SLLCE--REG-----ENDASRLKTEF--I--F--VQSAGDDEVLMGA--RPQE--E--VL--F--IKRVYVSL--NEET--LLLLKNLLCKQGSPLTQK

AFG3_YEAST DDMNLSGKLT--TLP--FT--ADIANACNE--LIAARHNDPYITIHFEO--IERVIAGLEKKTRV--SKEEKRSVAIHEAGE--VCGWFLKYADPLLKV
RCAL_YEAST DLKN-----RL--AL--P--FS--ADIANVCNE--LIAARSDEDAVKLNHFEO--IERVIGGVERKSKL--SPEEKVVVAIHEAGE--VCGWILKYADPLLKV
PARAPLEGIN FYSQ-----RL--EL--P--FS--ADIANICNE--LHAAREGHTSVHTLHFEY--VERVLGATAKKSKI--SKEEQKVAIFHESGR--LVGWMLEHTEAVMKV
YME1_YEAST PTII-----RG--PL--S--AELANLVNQ--VYACQKNVSVDMSHFEW--KDKILMGAERKTMV--TDAARKATAFHEAGE--IMAKYTNGATPLYKA
SPAST_HUMAN ELAQ-----L--RM--D--YS--SDLTALAKD--LGPIRELKEQVKN--MS--SEMRNIRLSDFTES--KKIKE--SVSPQTLQ--YIRWNKDFDQTTV--

AFG3_YEAST SIIPRCQCALGYAQYLPFDQYLISEEQFRHRMINALGGRVSEELHFF--SVTSGABDDFKKVTOMANAMVTSLGMSPKIGYLSFDQNDG--NFKV
RCAL_YEAST SIIPRCQCALGYAQYLPDIFLLTEOQLKDRMTMSLGGRVSEELHFF--SVTSGASDDFKKVTSMATAMVTELGMSDKIGMVNYQRD--DSDL
PARAPLEGIN SITPRTNALGYAQMLPRDQHLFTKEQLFERMCHALGGRASEALSFN--EVTSGAQDDLKRVTRIATSMVKQFGMAPGIGPISTFPAQEGLMGIG
YME1_YEAST TILPRGALGITFQLPMDKVDITKRECQARLDVCMGGKIAEELIYGKNDTTSGCCSDLOSATGTARAMVQYGMSSDDVGPVNLSENW-----
SPAST_HUMAN -----

AFG3_YEAST NKPFSSHKTARTIDLEVKSIVDDAHEACTELLTKNLDKVDLAKELLKKEAITREDMIRLLGPRP--FKERNEAFEKYLD-----PKSNTPE
RCAL_YEAST TKPFSDETCDIIDSEVYRIVQECHDECTELLKKEAEDVEKIAQVLLKKEVLTREDMIDLLGKRP--FPEBNDADFVKYLDYETEKIRKEEKNEKE
PARAPLEGIN RRPFSQGLQQMMDHEARLLVAKAYRHTKVLQDNLDKQLALANALLEKEVINYEDIEALIGPPPHGPKKMIAPQRWIDAQR--EKQDLGEEETEET
YME1_YEAST -ESWSKKIRDIADNEVIELLDSEERARELLTKKNVELERLAQGLIEYETLDAHEIEQVCKGEK--LDLKTSTNTVVEGPD--S--DERKDIGDEK
SPAST_HUMAN -----

AFG3_YEAST PEAPAAH-----
RCAL_YEAST NEPKPSTN-----
PARAPLEGIN QQPPLGGEEPWFK-----
YME1_YEAST PKIPTMLNA-----
SPAST_HUMAN -----

```

FIGURE 4B

Human: 1 [...] 459
 Mouse: 1 AGGCCGAGAGCGTCCGCGTCTTCCACAAGCAGGCCCTTCGAGTACATCTCCATTGCGCTGC 60
 Human: 460 AGGCCGAGCGCGTCCGAGTCTTCCACAACAGGCCCTTCGAGTACATCTCCATTGCGCTGC 519
 Mouse: 61 GCATCGACGAGGAAGAGAAAGCAGGACAGAAAGCAAGCTGTGGAATGGTATAAGAAAG 120
 Human: 520 GCATCGATGAGGATGAGAAAGCAGGACAGAAAGCAAGCTGTGGAATGGTATAAGAAAG 579
 Mouse: 121 GTATCGAAGAACTGGAAAAAGGAATCGCTGTTATAGTTACGGGCCAAGGTGAACAGTATG 180
 Human: 580 GTATTGAAGAACTGGAAAAAGGAATAGCTGTTATAGTTACAGGACAAGGTGAACAGTGTG 639
 Mouse: 181 AAAGAGCTAGACGCTCTCAAGCCAAAAATGATGACTAATTTAGTTATGGCCAAGGACCGCT 240
 Human: 640 AAAGAGCTAGACGCTCTCAAGCTAAAAATGATGACTAATTTGGTTATGGCCAAGGACCGCT 699
 Mouse: 241 TACAACCTTCTAGAGAAGCTGCAACCAAGTTTTCGAATTTTCCAAGTCACAGACGGAGCTCT 300
 Human: 700 TACAACCTTCTAGAGAAGATGCAACCAAGTTTTCGAATTTTCCAAGTCACAAACGGAGCTCT 759
 Mouse: 301 ATAACGAGAGTACTAAGCTGACATGCGCAATGGACATCTCCAGTCAGAAAGTGGAGCAG 360
 Human: 760 ATAATGACAGTACTAAGCTGGCATGCCGCAATGGACATCTCCAGTCAGAAAGTGGAGCTG 819
 Mouse: 361 TTCCGAAGAGGAAAGACCCCTTAACACATGCTAGTAATTCATTGCGCTCGATCAAAAACTG 420
 Human: 820 TTCCAAAAAGAAAGACCCCTTAACACACACTAGTAATTCAGTGCCTCGTTCAAAAAAG 879
 Mouse: 421 TCCGTAAAAAGTGGCTCCGCAAGGCTCTCCGGTCAACACAGGGCGCTAGTTGCAGTGGTT 480
 Human: 880 TTATGAAAACTGGATCTGCAGGCGCTTTCAGGCCACCATAGAGCACTAGTTACAGTGGTT 939
 Mouse: 481 TGTCCATGGTTTCTGGAGCAAGACCGGCACTGGTCTGCAGCTACCACACATAAGGGTA 540
 Human: 940 TATCCATGGTTTCTGGAGTGAAACAGGGATCTGGTCTGCTCTACCACTCATAGGGTA 999
 Mouse: 541 CTCCAAAAACCAATAGAACCAACAACTCTTACTCCCAACTGCAGTTTCGAAAAAAG 600
 Human: 1000 CTCCGAAAAACCAATAGGACAAATAACCTTCTACCCCTACAAGTCTACTCGTAAGAAAA 1059
 Mouse: 601 AAGACTTGAAAAATTTAGGAATGTGGACAGCAATCTTGCTAACCTTATAATGAATGAAA 660
 Human: 1060 AAGACTTGAAAAATTTAGGAATGTGGACAGCAACCTTGCTAACCTTATAATGAATGAAA 1119
 Mouse: 661 TTGTTGACAATGGGACAGCTGTAAAGTTGATGACATAGCCGGGAGGAGCTGGCAAGG 720
 Human: 1120 TTGTTGACAATGGGACAGCTGTAAATTTGATGATATAGCTGGTCAAGACTTGGCAAAAG 1179
 Mouse: 721 AAGCGCTGCAGGAGATTGTCATCTTCTCTCGGGCTGAGTTGTTACAGGGCTCA 780
 Human: 1180 AAGCATTGCAAGAAATTTGTTATCTTCTCTGAGGCTGAGTTGTTACAGGGCTTA 1239
 Mouse: 781 GAGCTCTGCTAGAGGCTTGTACTCTTCTGCGGCTGAGTAAACGAAAAACAATGCTGG 840
 Human: 1240 GAGCTCTGCGCAGAGGGCTGTTACTCTTGGTCCACCTGGGAATGGGAAGACAATGCTGG 1299
 Mouse: 841 CTAAAGCAGTAGCTGCAGAGTCTAATGCGACCTTTTCAACATAAGTGTCTGCCAGTTAA 900
 Human: 1300 CTAAAGCAGTAGCTGCAGAAATCGAATGCAACCTTCTTAAATAAGTGTGCAAGTTAA 1359
 Mouse: 901 CTTCAAAATATGTGGGAGAAGGAGAGAAATTTGTTGAGAGCTCTTTTGTGTGGCTCGAG 960
 Human: 1360 CTTCAAAATACGTGGGAGAAGGAGAGAAATTTGTTGAGGGCTCTTTTGTGTGGCTCGAG 1419
 Mouse: 961 AACTTCAACCATCTATAATTTTATAGATGAAGTTGACAGTCTTTTGTGTGAGAGACGGG 1020
 Human: 1420 AACTTCAACCTTCTATAATTTTATAGATGAAGTTGATAGCCTTTTGTGTGAAAGAAAGAG 1479
 Mouse: 1021 AAGGGGAGCAGCAGCTAGCAGACGGCTAAAGACGGAATTTTAAATAGAAATTTGACGGGG 1080
 Human: 1480 AAGGGGAGCAGCAGTGTCTAGTAGACGCTAAAACTGAATTTCTAATAGAAATTTGATGGTG 1539
 Mouse: 1081 TGCAATCTGCTGGAGATGACAGAGTACTTGAATGGGTGCAACTAACAGGCCCAAGAGC 1140
 Human: 1540 TACAGTCTGCTGGAGATGACAGAGTACTTGAATGGGTGCAACTAATAGGCCACAAGAGC 1599
 Mouse: 1141 TTGATGAAGCTGTTCTCAGGCGTTTCAATTAACGGGTATATGTCTCTTACCAATGAGG 1200
 Human: 1600 TTGATGAGGCTGTTCTCAGGCGTTTCAATTAACGGGTATATGTCTCTTACCAATGAGG 1659
 Mouse: 1201 AGACAAGACTCCTTCTGCTTAAAAACCTGTTGTGTAACAAGGAAGTCCACTGACCCAAA 1260
 Human: 1660 AGACAAGACTACTTTTGCTTAAAAATCTGTTATGTAACAAGGAAGTCCATTGACCCAAA 1719
 Mouse: 1261 AAGAAGCTCGCACAGCTTGTAGAAATGACCGATGGATACTCTGGAAGTGATCTGACCGCTT 1320
 Human: 1720 AAGAAGCTAGCACAACTTGTAGAAATGACTGATGGATACTCAGGAAGTGACCTAACAGCTT 1779
 Mouse: 1321 TGGCCAAGGATGCAAGCCCTGGGTCTATCCGAGAACTAAACCAAGAGGTGAAGAATA 1380
 Human: 1780 TGGCAAAAGATGCAAGCACTGGGTCTATCCGAGAACTAAACCAAGAGGTGAAGAATA 1839
 Mouse: 1381 TGTCTGCCAGTGAGATGAGAAATATTCGATTATCTGACTTCAAGAAATCTTAAAAAAGA 1440
 Human: 1840 TGTCTGCCAGTGAGATGAGAAATATTCGATTATCTGACTTCAAGAAATCTTAAAAAAGA 1899
 Mouse: 1441 TAAACCGCAGTGAGTCTCAGACCTTAGAAGCATACATACGCTGGAACAAGGATTTTG 1500
 Human: 1900 TAAACCGCAGCGTCAAGCCTCAAACTTTAGAAGCGTACATACGTTGGAACAAGGACTTTG 1959
 Mouse: 1501 GAGACACCACTGTTTAAAGGAAT 1523
 Human: 1960 GAGATACCACTGTTTAAAGGAAT 1982
 Human: 1983 [...] 3263
 Mouse: 1524 GGATGCCTCTGTGAGCCATAGAACATCGCACTTCACAGGAAACAAGAGCTTTGGCTACA 1583
 1584 GGAACCCAGACTTCGTTTACAGGAGCTTTAGAGTTTTCATTTTGTGCACCAAACTTGA 1643
 1644 AGAGGAACAAGAAGACAGACCTTAATAAATATGCAATATGAATGG 1689

FIGURE 5

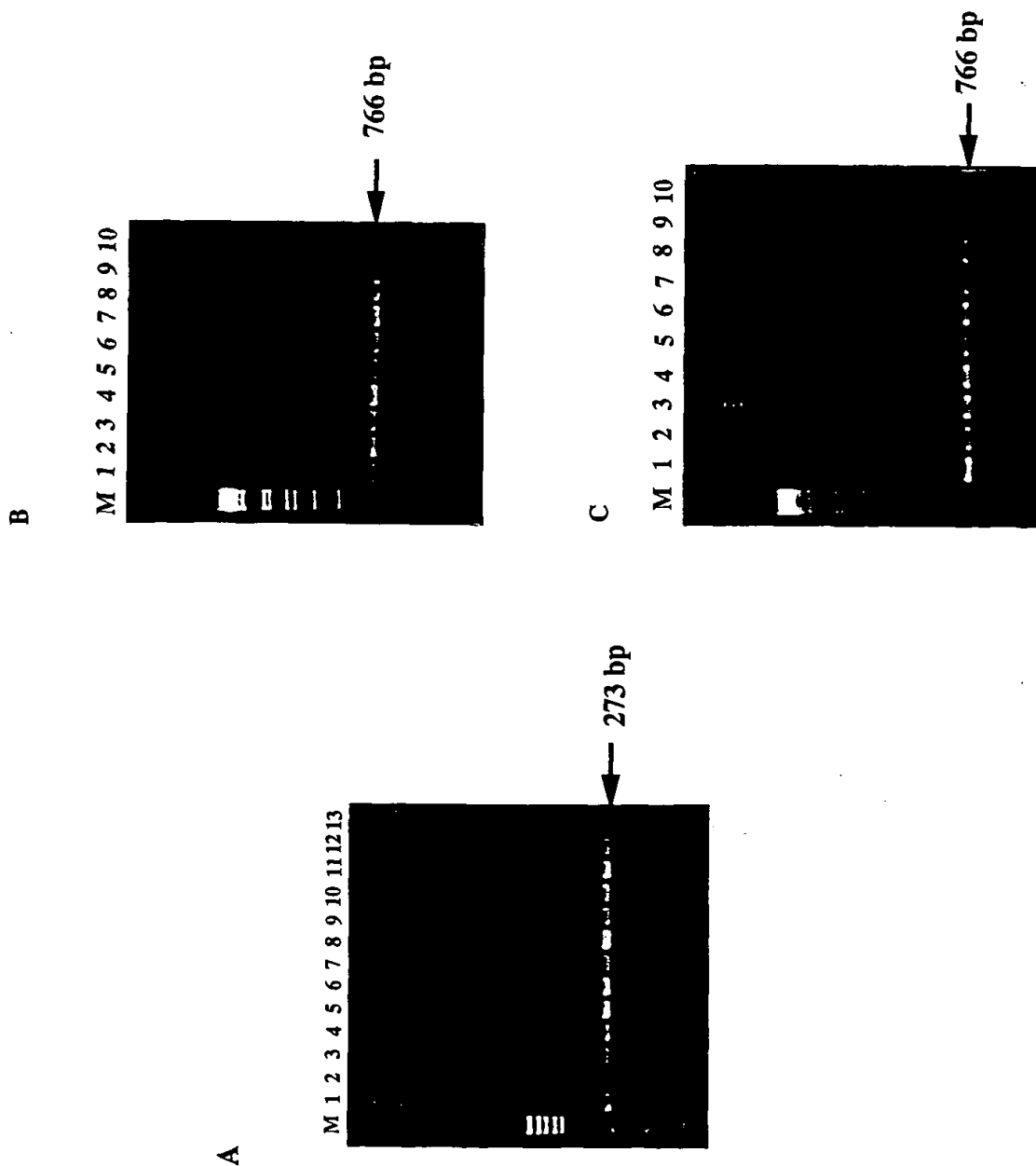


FIGURE 6

CABINET RECHERCHE
ORIGINAL